

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 56.6585 Seconds
(without alignments)
367.301 Million cell updates/sec

Title: US-09-675-208-5_COPY_510_610

Perfect score: 364
Sequence: 1 PDMDFOSTHPCAFHAASR.....LNKFTGVIGAFNCGGGWCR 101

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_minc:*
8: sp_organelle:*
9: sp-phage:*
10: sp_plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	564	100.0	784	10	Q92T62 cucumis sat
2	511	90.6	783	10	Q9FND9 arabidopsis
3	495	87.8	798	10	Q8VW6 pisum sativ
4	485	86.0	816	10	Q942T6 oryza sativ
5	409	72.5	857	10	Q9S8X0 phaseolus a
6	406	72.0	853	10	Q93XK2 pisum sativ
7	402	71.3	844	10	Q8RX87 arabidopsis
8	401	71.1	757	10	Q40077 hordeum vul
9	401	71.1	765	10	Q43408 brassica o1
10	400	70.9	773	10	Q9SCM1 arabidopsis
11	400	70.9	773	10	Q94A08 arabidopsis
12	392	69.5	386	10	Q9M442 cicer ariet
13	388	68.8	863	10	Q8RW08 stachys aff
14	379	67.2	779	10	Q9M4M7 persea amer
15	371	65.8	807	10	Q9ST04 arabidopsis
16	341	60.5	125	10	Q42099 arabidopsis

17	213	37.8	1170	10	Q9LFE7	Q91FZ7 arabidopsis
18	183	32.4	649	17	Q97U94	Q97U94 sulfolobus
19	167	29.6	674	17	Q96XG2	Q96XG2 sulfolobus
20	147.5	26.2	702	2	Q93CM6	Q93CM6 bifidobacte
21	115	20.4	357	10	Q93A66	Q93A66 cicer ariet
22	104.5	18.5	371	10	Q04607	Q04607 arabidopsis
23	80	14.2	283	10	Q9FWM2	Q9FWM2 arabidopsis
24	75	13.3	576	16	Q9ZCM8	Q9ZCM8 rickettsia
25	73	12.9	348	8	Q94W55	Q94W55 coryphopter
26	72	12.8	840	10	Q9LW46	Q9LW46 arabidopsis
27	71	12.6	273	16	Q9PN62	Q9PN62 campylobact
28	71	12.6	382	17	Q970S3	Q970S3 sulfolobus
29	71	12.6	638	5	Q9VYU8	Q9VYU8 drosophila
30	71	12.6	959	12	Q55716	Q55716 chilo iride
31	70.5	12.5	212	2	Q49799	Q49799 mycobacteri
32	70.5	12.5	236	9	Q64059	Q64059 bacterioph
33	70.5	12.5	236	16	Q91962	Q91962 bacillus su
34	70.5	12.5	294	16	Q9S377	Q9S377 mycobacteri
35	69	12.2	615	4	Q9NUX3	Q9NUX3 homo sapien
36	69	12.2	615	4	Q8TAP0	Q8TAP0 homo sapien
37	69	12.2	623	3	Q9P938	Q9P938 plichia stip
38	69	12.2	2023	12	Q91632	Q91632 cherry tree
39	68.5	12.1	183	16	Q911Y7	Q911Y7 pseudomonas
40	68.5	12.1	216	4	Q9BT41	Q9BT41 homo sapien
41	68.5	12.1	219	4	Q9BVW4	Q9BVW4 homo sapien
42	68	12.1	329	17	Q8TWT1	Q8TWT1 methanopyru
43	68	12.1	2026	12	Q9DVT1	Q9DVT1 cherry tree
44	67	11.9	292	16	Q9AB15	Q9AB15 caulobacter
45	67	11.9	373	5	Q26960	Q26960 trypanosoma

ALIGNMENTS

RESULT 1	Q92T62	PRELIMINARY:	PRT:	784 AA.
ID	Q92T62			
AC	Q92T62			
DT	01-MAY-1999 (TREMBLrel. 10, created)			
DT	01-MAY-1999 (TREMBLrel. 10, last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, last annotation update)			
DE	Raffinose synthase (EC 2.4.1.82).			
GN	RFS.			
OS	Cucumis sativus (Cucumber).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid1; Cucurbitales; Cucurbitaceae; Cucumis.			
OX	NCBI_TaxID=3659;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAF;			
RA	Ohsumi C., Nozaki J., Kida T.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF073744; AAD02832.1; -			
KW	Glycosyltransferase; Transferase.			
SQ	SEQUENCE 784 AA: 86920 MW: 38064491P0908933 CRC64;			
Query Match	100.0%:	Score 564:	DB 10:	Length 784:
Best Local Similarity	100.0%:	Pred. No. 17e-54:		
Matches 101:	Conservative	0:	Mismatches	0:
			Indels	0:
			Gaps	0:
OY	1 PDMDFOSTHPCAFHAASRAISGGPIYSDVGKHNFDLKKLVPGSILRSEYALP 60			
DB	510 PDMDFOSTHPCAFHAASRAISGGPIYSDVGKHNFDLKKLVPGSILRSEYALP 569			
OY	61 TRDCLFEDPLHNGFTMLKIWNLNKFTGVIGAFNCGGGWCR 101			
DB	570 TRDCLFEDPLHNGFTMLKIWNLNKFTGVIGAFNCGGGWCR 610			
RESULT 2	Q9FND9	PRELIMINARY:	PRT:	783 AA.
ID	Q9FND9			

```
AC Q9FND9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Raffinose synthase protein.
GN MPO12.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
RX NCBL_TaxID=3702;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLMBIA;
RC MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones."
RL DNA Res. 4:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB006702; BAB1595.1; -
DR EMBL; AY062781; BAB32859.1; -
DR EMBL; AY081645; AAM10207.1; -
SQ SEQUENCE 783 AA; 86237 MW; 3C37DD1D7871888AC CRC64;

Query Match
Best Local Similarity 90.6%; Score 511; DB 10; Length 783;
Matches 90; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 PWDMMFQSTHPCAFPAFAASRAISGPIYVSDSYGKHFLLKLVLPDGSILRSEYALP 60
DB 515 PWDMMFQSTHPCAFPAFAASRAISGPIYVSDSYGKHFLLKLVLPDGSILRSEYALP 574
QY 61 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCGCGGMC 101
DB 575 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCGCGGMC 615

RESULT 3
O8VYNG PRELIMINARY; PRT; 798 AA.
AC O8VYNG;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Raffinose synthase (EC 2.4.1.82).
GN RFS.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
RX NCBL_TaxID=3888;
OX [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=CV. WUNDER VON KELVEDON; TISSUE=SEED;
RA Peterbauer T., Mach L., Mucha J., Richter A.;
RT "Molecular characterization of raffinose synthase from pea (Pisum
RT sativum L.) seeds."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0426475; CAD20127.2; -
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 798 AA; 88717 MW; 8D3F3ED5BF8617B7 CRC64;

Query Match
Best Local Similarity 87.8%; Score 495; DB 10; Length 798;
Matches 85; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 PWDMMFQSTHPCAFPAFAASRAISGPIYVSDSYGKHFLLKLVLPDGSILRSEYALP 60
DB 528 PWDMMFQSTHPCAFPAFAASRAISGPIYVSDSYGKHFLLKLVLPDGSILRSEYALP 587
QY 61 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCGCGGMC 100
DB 588 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCGCGGMC 627

RESULT 4
Q942T6 PRELIMINARY; PRT; 816 AA.
AC Q942T6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Putative raffinose synthase.
GN P0583G08.2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBL_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0583G08."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003282; BAB64768.1; -
SQ SEQUENCE 816 AA; 89588 MW; B316EDF3566C5178 CRC64;

Query Match
Best Local Similarity 86.0%; Score 485; DB 10; Length 816;
Matches 83; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 PWDMMFQSTHPCAFPAFAASRAISGPIYVSDSYGKHFLLKLVLPDGSILRSEYALP 60
DB 546 PWDMMFQSTHPCAFPAFAASRAISGPIYVSDSYGKHFLLKLVLPDGSILRSEYALP 605
QY 61 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCGCGGMC 101
DB 606 TRDCLFADPLHDGKTMKIMNLKFTGVIGAFNCGCGGMSR 646

RESULT 5
Q9SBZ0 PRELIMINARY; PRT; 857 AA.
AC Q9SBZ0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Galactinol-raffinose galactosyltransferase (EC 2.4.1.67).
OS Phaseolus angularis (Adzuki bean) (Vigna angularis).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX NCBL_TaxID=3914;
RN [1]
RP SEQUENCE FROM N.A.
```


overlapped

RP SEQUENCE FROM N.A.
 RA MEDLINE-20117502; PubMed-10652123;
 RC Peterbauer T., Mucha J., Mayer U., Popp M., Gloessel J., Richter A.;
 RT "Stachyose synthesis in seeds of adzuki bean (Vigna angularis):
 RL molecular cloning and functional expression of stachyose synthase";
 RL Plant J. 20:309-318(1999).
 DR EMBL: Y19024; CAB64363.1; -;
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 857 AA; 94949 MW; 85248C4B81165679 CRC64;

Query Match 72.5%; Score 409; DB 10; Length 857;
 Best Local Similarity 70.7%; Pred. No. 4.4e-37;
 Matches 70; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

OY 1 PDMDFOSTHPCAAFHNSRAISGPIYSDSVGKHNFDLKKLVLPDGSILRSEYALP 60
 DB 586 PDMDFOSHCAKFAHNSRAISGPIYSDSVGSHDFDLKKLVLPDGSILRSEYALP 645
 OY 61 TRDCLFEDPLHNGETMLKIMNFKFTGVIGAFNCGGCM 99
 DB 646 TRDCLFERNLFDQKTVLKIMNFKYGVIGAFNCGGCM 684

RESULT 6

OY3XK2 PRELIMINARY; PRT; 853 AA.

AC OY3XK2;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Stachyose synthase (EC 2.4.1.67).
 CN STS1.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 OX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. MUNDER VON KEVEDON; TISSUE=SEED;
 RA Peterbauer T.;
 RT "Molecular cloning of a cDNA encoding for stachyose synthase from pea
 seeds";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ311087; CAC38094.1; -;
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 853 AA; 95890 MW; CB10F18CB037B38C CRC64;

Query Match 72.0%; Score 406; DB 10; Length 853;
 Best Local Similarity 71.7%; Pred. No. 9.4e-37;
 Matches 71; Conservative 11; Mismatches 17; Indels 0; Gaps 0;

OY 1 PDMDFOSTHPCAAFHNSRAISGPIYSDSVGKHNFDLKKLVLPDGSILRSEYALP 60
 DB 583 PDMDFOSHCAKFAHNSRAISGPIYSDSVGSHDFDLKKLVLPDGSILRSEYALP 642
 OY 61 TRDCLFEDPLHNGETMLKIMNFKFTGVIGAFNCGGCM 99
 DB 643 TRDCLFERNLFDQKTVLKIMNFKYGVIGAFNCGGCM 681

RESULT 7

OY3X87 PRELIMINARY; PRT; 844 AA.

AC OY3X87;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE AT5G20250/F5024_140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamita A., Karlin-Neumann G.,
 RA Kawai S., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Tortum M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY090237; AAL90901.1; -;
 SQ SEQUENCE 844 AA; 94215 MW; 2AC8AB0E4A3F8056 CRC64;

Query Match 71.3%; Score 402; DB 10; Length 844;
 Best Local Similarity 68.7%; Pred. No. 2.6e-36;
 Matches 68; Conservative 14; Mismatches 17; Indels 0; Gaps 0;

OY 1 PDMDFOSTHPCAAFHNSRAISGPIYSDSVGKHNFDLKKLVLPDGSILRSEYALP 60
 DB 571 PDMDFHSHVPAEYHNSRAISGPIYSDSVGKHNFDLKKLVLPDGSILRSEYALP 630
 OY 61 TRDCLFEDPLHNGETMLKIMNFKFTGVIGAFNCGGCM 99
 DB 631 TRDCLFADPARGVSLKIMNFKYGVIGAFNCGGCM 669

RESULT 8

OY40077 PRELIMINARY; PRT; 757 AA.

AC OY40077;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Seed imbibition protein.
 CN SIPI.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HIMALAYA;
 RA Heck G.R., Dorsett C., Ho T.-H.D.;
 RT "Cloning and characterization of a gene, SIPI, associated with seed
 imbibition in barley";
 RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
 DR EMBL: M77475; AAA32975.1; -;
 SQ SEQUENCE 757 AA; 82133 MW; EA9E7B771AABBCF CRC64;

Query Match 71.1%; Score 401; DB 10; Length 757;
 Best Local Similarity 67.3%; Pred. No. 2.9e-36;
 Matches 68; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

OY 1 PDMDFOSTHPCAAFHNSRAISGPIYSDSVGKHNFDLKKLVLPDGSILRSEYALP 60
 DB 483 PDMDFHSHVPAEYHNSRAISGPIYSDSVGKHNFDLKKLVLPDGSILRSEYALP 642
 OY 61 TRDCLFEDPLHNGETMLKIMNFKFTGVIGAFNCGGCM 101
 DB 543 TRDCLFADPARGVSLKIMNFKYGVIGAFNCGGCM 683

RESULT 9

OY43408 PRELIMINARY; PRT; 765 AA.

AC OY43408;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Putative imbibition protein.
OS Brassica oleracea (Caulliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ALPHA BALOMA; TISSUE=ROOT TIPS;
RA Fujikura Y., Karsen C.K.;
RT "Caulliflower cDNA encoding a putative imbibition protein."
DR EMBL: XJ9330; CAAS5893.1; -
SQ SEQUENCE 765 AA; 84084 MW; 73398603048E9B58 CRC64;
Query Match 71.1%; Score 401; DB 10; Length 765;
Best Local Similarity 66.3%; Pred. No. 3e-36;
Matches 67; Conservative 13; Mismatches 21; Indels 0; Gaps 0;
QY 1 PWDMMFQSTHPCAFHAAASRAISGPIYVSDSVGKHNFDLKLKLYLPDGSILRSEYALP 60
DB 476 PWDMMFHSILHPTAEYHAAARAVGCAIYVSDKPGNNHFDLRLKLYLPDGSVLRALPGRP 535
QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCQGGGMC 101
DB 536 TRDCLFADPARDDTSLKTIWNNMKFTGIVGVFCQAGCMCK 576
RESULT 10
Q9SCM1 PRELIMINARY; PRT; 773 AA.
ID Q9SCM1
AC Q9SCM1;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE Imbibition protein homolog.
GN TBH10.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansover W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133248; CAB6109.1; -
SQ SEQUENCE 773 AA; 85143 MW; 0852F9E67952C8D3 CRC64;
Query Match 70.9%; Score 400; DB 10; Length 773;
Best Local Similarity 66.3%; Pred. No. 3.9e-36;
Matches 67; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
QY 1 PWDMMFQSTHPCAFHAAASRAISGPIYVSDSVGKHNFDLKLKLYLPDGSILRSEYALP 60
DB 478 PWDMMFHSILHPTAEYHAAARAVGCAIYVSDKPGNNHFDLRLKLYLPDGSVLRALPGRP 537
QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCQGGGMC 101
DB 538 TRDCLFADPARDDTSLKTIWNNMKFTGIVGVFCQAGCMCK 578
RESULT 11
Q94A08 PRELIMINARY; PRT; 773 AA.
ID Q94A08
AC Q94A08;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Putative imbibition protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onders C.S., Quach H.L.,
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carinuci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koessma E., Lam B.,
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou R., Seki M., Shino P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene TBH10.120/At3g57520 (GI:5706423)."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY050772; AAK92707.1; -
SQ SEQUENCE 773 AA; 85171 MW; 18BDF9E67952C801 CRC64;
Query Match 70.9%; Score 400; DB 10; Length 773;
Best Local Similarity 66.3%; Pred. No. 3.9e-36;
Matches 67; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
QY 1 PWDMMFQSTHPCAFHAAASRAISGPIYVSDSVGKHNFDLKLKLYLPDGSILRSEYALP 60
DB 478 PWDMMFHSILHPTAEYHAAARAVGCAIYVSDKPGNNHFDLRLKLYLPDGSVLRALPGRP 537
QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCQGGGMC 101
DB 538 TRDCLFADPARDDTSLKTIWNNMKFTGIVGVFCQAGCMCK 578
RESULT 12
Q9M442 PRELIMINARY; PRT; 386 AA.
ID Q9M442
AC Q9M442;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE Putative imbibition protein (Fragment).
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.
OX NCBI_TaxID=3827;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CASTELLANA; TISSUE=ETIOLATED OSMOTIC STRESSED EPICOTYLS;
RA Dopico B., Romo S., Labrador E.;
RT "A putative imbibition protein is expressed in chickpea epicotyls."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ271668; CAB71135.1; -
FT NON_TER
SQ SEQUENCE 386 AA; 42180 MW; 357AF9FBC8D71650 CRC64;
Query Match 69.5%; Score 392; DB 10; Length 386;
Best Local Similarity 67.3%; Pred. No. 1.3e-35;
Matches 68; Conservative 11; Mismatches 22; Indels 0; Gaps 0;
QY 1 PWDMMFQSTHPCAFHAAASRAISGPIYVSDSVGKHNFDLKLKLYLPDGSILRSEYALP 60
DB 90 PWDMMFHSILHPTAEYHAAARAVGCAIYVSDKPGNNHFDLKLKLYLPDGSVLRALPGRP 149
QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCQGGGMC 101
DB 150 TRDCLFVDPARDTSLKTIWNNMKFTGIVGVFCQAGCMCK 190
RESULT 13
Q8RW08

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 28.9451 Seconds
(Without alignments)
335.448 Million cell updates/sec

Title: US-09-675-208-5_COPY_510_610

Perfect score: 564

Sequence: 1 PDMDFQSTHPCAAFHNASR.....LNKFTGVIGAFNCGGCMCR 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_73:**

1: p1r1:**
2: p1r2:**
3: p1r3:**
4: p1r4:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	401	71.1	757	2 S27762	S1p1 protein - bar
2	401	71.1	765	2 S45033	probable imbibition
3	400	70.9	773	2 T46188	Imbibition protein
4	371	65.8	807	2 C85025	hypothetical prote
5	213	37.8	1170	2 C96599	protein F20N2.14
6	183	32.4	649	2 D90496	hypothetical prote
7	115	20.4	357	2 T09530	probable seed inh
8	104.5	18.5	371	2 T01737	hypothetical prote
9	76.5	13.6	165	2 F70392	hypothetical prote
10	75	13.3	576	1 B71676	probable ABC trans
11	71	12.6	273	2 E81330	ribonucleoside-dip
12	71	12.6	959	2 T03053	hypothetical prote
13	70.5	12.5	212	2 S72873	hypothetical prote
14	70.5	12.5	236	2 T12810	hypothetical prote
15	70.5	12.5	294	2 B87073	conserved hypotet
16	69.5	12.3	544	2 A72459	probable glutamyl-
17	68.5	12.1	183	2 G83379	probable fibrillar
18	67	11.9	292	2 C87275	hydrolyase, carbon-
19	66.5	11.8	149	2 A71331	conserved hypotet
20	66	11.7	1291	2 T40631	non-proteolytic bo
21	65.5	11.6	336	2 H84767	probable peroxidase
22	64.5	11.4	172	1 C71128	hypothetical prote
23	64.5	11.4	211	2 T49930	hypothetical prote
24	64.5	11.4	358	2 S57176	hypothetical prote
25	64.5	11.4	498	2 T23525	hypothetical prote
26	64.5	11.4	626	2 S53871	Pmel 17 protein -
27	64.5	11.4	667	2 A48579	trophozoite surfac
28	64	11.3	105	1 H64085	g1pE protein homol
29	64	11.3	406	2 H71694	proline/betaine tr

30	64	11.3	486	2 T31294	hypothetical prote
31	64	11.3	674	2 G72204	hypothetical prote
32	64	11.3	1541	2 AG2474	heterocyst glycoli
33	63.5	11.3	402	1 MNXRPC	nonstructural prot
34	63.5	11.3	418	2 A85168	IS1167, transposas
35	63.5	11.3	418	2 B95197	IS1167, transposas
36	63.5	11.3	418	2 E95146	IS1167, transposas
37	63.5	11.3	418	2 H95208	IS1167, transposas
38	63.5	11.3	429	2 C71075	probable cytochrom
39	63.5	11.3	819	2 T10355	hypothetical prote
40	63	11.2	108	2 T31283	naphthalene 1,2-di
41	63	11.2	252	2 F82640	hypothetical prote
42	63	11.2	355	2 C95843	probable DNA ligase
43	63	11.2	884	2 S53396	telomerase catalyt
44	63	11.2	982	1 S16292	nitrate reductase
45	62.5	11.1	165	2 H71649	hypothetical prote

ALIGNMENTS

102(6)

RESULT 1

S27762

S1p1 protein - barley

C:Species: Hordeum vulgare (barley)

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997

C:Accession: S27762

R:Heck, G.R.; Dorsett, C.; Ho, T.H.

Submitted to the EMBL Data Library, February 1992

A:Reference number: S27762

A:Accession: S27762

A:Molecule type: DNA

A:Residues: 1-757 <HEC>

A:Cross-references: EMBL:M77475; NID:q167099; PID:q167100

A:Genetics:

A:Introns: 64/2; 151/1; 181/2; 197/3; 489/3; 557/2

Query Match 71.1% Score 401; DB 2; Length 757;

Best Local Similarity 67.3%; Pred. No. 8e-35; Mismatches 22; Indels 0; Gaps 0;

Matches 68; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

Qy 1 PDMDFQSTHPCAAFHNASRAISGPIYSDSVGKHNFDLKLVLPDGSILRSEYVALP 60

Db 483 PDMDFHSIHPAEYHNGARATGCGPIYSDKRGKHNFDLKLVLPDGSILRQLPGRP 542

Qy 61 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGCMCR 101

Db 543 TRDCLFSDPARDGASLTKIWNKAGVGVNCGGCMCR 583

RESULT 2

S45033

probable imbibition protein - wild cabbage

C:Species: Brassica oleracea (wild cabbage)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997

C:Accession: S45033

R:Fujikura, Y.; Karlsen, C.K.

Submitted to the EMBL Data Library, May 1994

A:Description: Cauliflower cDNA encoding a putative imbibition protein.

A:Reference number: S45033

A:Accession: S45033

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-765 <FUD>

A:Cross-references: EMBL:X79330; NID:q488786; PID:q488787

Query Match 71.1% Score 401; DB 2; Length 765;

Best Local Similarity 66.3%; Pred. No. 8.1e-35; Mismatches 21; Indels 0; Gaps 0;

Matches 67; Conservative 13; Mismatches 21; Indels 0; Gaps 0;

Qy 1 PDMDFQSTHPCAAFHNASRAISGPIYSDSVGKHNFDLKLVLPDGSILRSEYVALP 60

Db 483 PDMDFHSIHPAEYHNGARATGCGPIYSDKRGKHNFDLKLVLPDGSILRQLPGRP 542

Qy 61 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGCMCR 101

Db 543 TRDCLFSDPARDGASLTKIWNKAGVGVNCGGCMCR 583

Dd 476 PDMFMHSLHTAETAEYHAARAVGGCAIIVSDKPCNNHFDDLRLKVLVLPDGSYLRRLLPGRP 553

Qy 61 TRDCLEDPDLNGEFTMLKIYNLKNFTGVIAGFNQQSGGMCR 101
||||| : : |||||:|||||:||||| |||:
Db 536 TRDCLEFADPARDGSTLKIKYMNKNFTGVIAGFNQQSGGMCK 576

RESULT 3

Imbibition protein homolog - *Arabidopsis thaliana*
 N:Alternate names: protein T8H10.120
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T46188
 R:Benes, V.; Rechmann, S.; Borikova, D.; Anstorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K. I. G.
 submitted to the Protein Sequence Database, January 2000
 ;Reference number: 223014

A;Note: T8H10.120

Db 478 PDWDMFHSLHPTAEYHAAARAVGGCAIYVSDKPGNHNFLLRKLVLPGGSVLRAKLPGRP 53

```

Oy      61 TRDCLFEDPLHNGETMLKIWNLNKFVGIAFNCQGGGWCR 101
         ||||| | : : ||||| : ||||| : ||||| : ||||| :
Db      538 TRDCLFADPARDGISLLKIWNMNFVTGIVGVFNCQGAGWCK 578

```

RESULT 4

hypothetical protein AT4g01970 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001
C:Accession: 885025
C:anonymous: The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198

Query Match	65.8%	Score 371;	DB 2;	Length 807;
Best Local Similarity	67.3%	Pred. NO. 1.4e-31;		
Matches 68; Conservative	13;	Mismatches 18;	Indels 2;	Gaps 1

Qy 1 PDMDMFQSTHPCAAFAAASRAISGGPIYVSDSYGK--HNFDLKLVLPDGSILRSEYA 58
||||||| | : ||||| | : : || ||||| : ||
Db 530 PDMDMFQSDHYCAEYHAASRAICGGPIYVLSDLHGKASHNFDLIKLAFFDGTIPRCVHYA 5899

```

QY      59  LPTRDCLFEDPLHNGETMLKIWNLKNFTGIVIGAFNCQGGGW  99
        |||||::|:|::|::|::|::|::|::|::|::|::|::|::|
DB      590  LPTRDSLAKNPFLFDKESILKIFENFKFGVIGTFNCQAGAW  630

```

RESULT 5
C96599

Protein F20N2.4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence-revision 02-Mar-2001 #text-change 31-Mar-2001
C:Accession: C965999
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar
ansen, N.F.; Hughes, B.; Huizar, L.
ature 408, 816-820, 2000

Query Match	37.8%	Score 213	DB 2	Length 1170
Best Local Similarity	59.1%	Pred. No. 1,9e-14		
Matches	39	Conservative	12	Mismatches 15, Indels 0, Gaps 0

QY 1 PDWDMQSTHPCAAIFHAASRAISGGPIYVSDSVGKHNFDLKLVLPGDISLRSEYVALP 60

QY	61	TRDCLF	66
		:::	
Db	558	TRELTV	563

RESULT 6

hypothetical protein SS03127 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: D904456
R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Azeazy, M.J.; Chae, J.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi Ngoc, H.P.; Redder, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
Submitted to Genbank, April 2001
#Description: Sulfolobus solfataricus complete genome.

Query Match	32.4%	Score 183;	DB 2;	length 649;
Best Local Similarity	41.6%;	Pred No. 1.5e-11;		
Matches 42;	Conservative 14;	Mismatches 37;	Indels 8;	Gaps 3

```

QY      1 PDWMFGSTHPCAAFAAASRAISGGPIVSDSGKH---NFDLLKLVLPDGSILRSEY 56
      ||:||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB     455 PDYDMFMSYDPAKYVHLVARVSGGPIITYD---RHPERTNIELLRMAVLPGCEVIRDE 511

```

QY 57 YALPTRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCQGG 97
 ||| ||| ||| :||| ||| |
 Db 512 PALITTEDLLEFKDPLRE-RVLLKLKGVKGYNAIAFNLNSG 551

RESULT 7

[illegible]

```

C:Genetics:
A:Gene: eq_1075

Query Match      13.6%: Score 76.5; DB 2; Length 165;
Best Local Similarity 29.5%: Pred. No. 0.79;
Matches 28; Conservative 12; Mismatches 36; Indels 19; Gaps 6;

OY 1 PDMFMFQSTHPCAFHAASRAISGPIYSDS--VGKINFDLKLVL---LPDGSIL--R 53
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 82 PKWIMFAARLPGRKLRHPK----SGGKYRDEAVIVGEGNVELEPNFISFKYSDGRIINFR 137
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 54 SEYVALPTPRDCLFEDPLCHNGETMLKIMWLNKRTGV 88
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 IEEY----RYLRKYSPKYGPSPFRSEYW---FIGV 164
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
B71676
Probable ABC transporter msbA2 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Nov-2000
C:Accession: B71676
R.Andersson, S.G.-E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: B71676
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-576 <AND>
A:Cross-references: GB:AD235272; GB:AD235269; NID:g3861033; PIDN:CAAI5132.1; PID:g386
A:Experimental source: Strain Madrid E
C:Genetics:
A:Gene: msbA2; RP696
C:Superfamily: Rickettsia prowazekii probable ABC transporter msbA2; ATP-binding class
C:Keywords: ATP
F:354-548/Domain: ATP-binding cassette homology <ABC>

Query Match      13.3%: Score 75; DB 1; Length 576;
Best Local Similarity 27.0%: Pred. No. 5;
Matches 27; Conservative 14; Mismatches 25; Indels 34; Gaps 5;

OY 15 FHASRAISGPIYSDSVGKHNDLKLKLYLP-----DGSILRSEYALPTR 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 282 YVAILAGVSSGCI-----FELLSEMHLPPTALEIRITIIDKTSIVHNYYALNNS 331
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 63 DCL---FE--DPLHNGETMLKI-----WNLNKTGYIG 90
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 332 DAISIEFKNVDEFTYNSRPNLKVINMSLKNKFGIVG 371
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
B81330
Probable peptidylase protein Cj1235 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: B81330
R.Parkhill, J.; Wren, B.W.; Mungall, K.; Kellie, J.M.; Churcher, C.; Basham, D.; Chillingworth, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: AB1250; MUID:20150912; PMID:10688204
A:Accession: B81330
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-273 <PAR>
A:Cross-references: GB:AL139077; GB:AL11168; NID:g6968444; PIDN:CAE73489.1; PID:g696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1235

Query Match      12.6%: Score 71; DB 2; Length 273;
Best Local Similarity 26.0%: Pred. No. 5.6;

```


Best Local Similarity 34.8%; Pred. No. 6.9;
Matches 23; Conservative 8; Mismatches 18; Indels 17; Gaps 5;

0Y 41 LKKLVLDGSIRSEYYALPTRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFN-----C 94

```
Db      204 LRLIYLADSQLLD---YA-PDRDEL---LRFCKTLMAIWRTIQFSGETGDFRPNPSRLC 255
```

QY 95 QGGCWC 100

Db 256 ---GWC 258

Search completed: March 27, 2003, 10:06:54
Job time : 30.9451 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; search time 16.628 seconds

(without alignments)
251.930 million cell updates/sec

Title: US-09-675-208-5_COPY_510_610

Perfect score: 564
Sequence: 1 PDMDQSTHPCAFHAFHAAASR.....LNKFTGVIGAFNCQGGWCR 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76.5	13.6	165	YA75_AQUAE	067169 aquifex aeo
2	69.5	12.3	544	SYE_AERPE	0949h1 aeropyrum p
3	67	11.9	258	FLA3_PYRKO	0942w9 pyrococcus
4	66.5	11.8	149	MRA2_TREPA	083398 treponema p
5	64.5	11.4	358	YJ9L_YEAST	P47175 saccharomyc
6	64.5	11.4	626	PM17_MOUSE	060696 mus musculus
7	64.5	11.4	667	TS11_GIALA	003385 giardia lam
8	64	11.3	105	GLPE_HAEIN	P44819 haemophilus
9	63.5	11.3	402	VN34_ROTPO	P27586 porcine rot
10	63.5	11.3	819	Y083_NPVOP	010336 oryza pseu
11	63.5	11.2	721	BBS2_RAT	099mh9 ratulus norv
12	63	11.2	884	TERT_YEAST	006163 saccharomyc
13	63	11.2	982	NIA_NEUCR	P08619 neurospora
14	62.5	11.1	165	Y872_RICPR	092694 rickettsia
15	62.5	11.1	220	AP12_SOUTU	043646 solanum tub
16	62.5	11.1	263	YC56_PORPU	P51308 porphyra pu
17	62.5	11.1	266	SUHB_BUCAI	P57372 buchiera ap
18	62.5	11.1	481	HTR3_HALNI	09h000 halobacteri
19	62.5	11.1	481	HTR3_HALSA	048316 halobacteri
20	62	11.0	617	VAAT_HUMAN	P38606 homo sapien
21	62	11.0	993	ROXN_HUMAN	Q94972 homo sapien
22	62	11.0	1264	YGN3_YEAST	P53125 saccharomyc
23	61.5	10.9	295	DAP1_BACHD	09K632 bacillus ha
24	61.5	10.9	346	CCPR_PSEAE	P14532 pseudomonas
25	61	10.8	590	VPP_BPP2	P25479 bacteriophy
26	60.5	10.7	249	RNPL_YEAST	P32385 saccharomyc
27	60.5	10.7	293	MOTV_VIBPA	P46233 vibrio para
28	60.5	10.7	340	CELF_VZVD	P09261 varicella-z
29	60.5	10.7	715	PAPI_XENLA	P51004 xenopus lae
30	60.5	10.7	1044	IFZP_PYRHO	058822 pyrococcus
31	60	10.6	210	RHOD_HUMAN	000212 homo sapien
32	60	10.6	235	SOML_GADMO	P21919 gadus morhu
33	60	10.6	294	DAPA_RICCN	092125 rickettsia

34	60	10.6	294	DAPA_RICMO	Q9akg3 rickettsia
35	60	10.6	294	DAPA_RICRI	Q9akj9 rickettsia
36	60	10.6	765	TREG_PHYBL	P20409 p anthranil
37	59.5	10.5	188	API8_SOLRU	P16348 solanum tub
38	59.5	10.5	220	API8_SOLTU	P17979 solanum tub
39	59.5	10.5	358	T2H2_HAEPA	P36433 haemophilus
40	59.5	10.5	423	MTN4_NEILA	P50182 neisseria 1
41	59.5	10.5	525	Y487_TREPA	083500 treponema p
42	59.5	10.5	529	Y119_NPVOP	010358 oryza pseu
43	59.5	10.5	717	YM4_CABEL	P34520 caenorhabdl
44	59	10.5	377	AMPC_ECOLI	P00811 escherichia
45	59	10.5	390	ARP2_SCHPO	Q9uu11 schizosach

ALIGNMENTS

RESULT 1	ID	YA75_AQUAE	STANDARD:	PRT:	165 AA.
AC	067169;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Hypothetical protein AQ_1075.				
GN	AQ_1075.				
OS	Aquifex aeolicus.				
OC	Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;				
OX	Aquifex.				
NCBI_TaxID=63363;					
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=VF5;				
RX	MEDLINE=98196666; PubMed=9537320;				
RA	Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,				
RA	Graham D.E., Overbeek R., Sneed M.A., Aujeay M., Huber R.,				
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;				
RT	"The complete genome of the hyperthermophilic bacterium Aquifex				
RT	aeolicus.";				
RL	Nature 392:353-358(1998).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement. (See http://www.isb-sdb.ch/announce/				
CC	or send an email to license@isb-sdb.ch).				
CC	-----				
DR	EMBL: AEO00721; AAC07132.1; -				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 165 AA; 19532 MW; 027AB8C916795818 CRC64;				
Query Match	13.6%; Score 76.5; DB 1; Length 165;				
Best Local Similarity	29.5%; Pred. No. 0.22;				
Matches	28; Conservative 12; Mismatches 36; Indels 19; Gaps 6;				
QY	1 PDMDQSTHPCAFHAFHAAASRISGCPYYSDS--VGKHNFDLLKIV---LPGCSIL--R 53				
DB	82 PKWIMFAARLPNCKLHPK---SGGKYVRDVAIVGEGNLEPNFTSPKYSQGRILNFR 137				
QY	54 SEVYALPTRDCLFEDPLNGETMLKIWNLNKFTGV 88				
DB	138 IERY-----RYLRKSPKRYGSPFSFEW---FIGV 164				
RESULT 2					
SYE_AERPE					
ID	SYE_AERPE	STANDARD:	PRT:	544 AA.	
AC	0949h1;				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				

DE glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
DE (GURS).
GN GLTX OR APE2317.
OS Aeropyrum pernix.
OC Archaea: Crenarchaeota: Thermoprotei: Desulfurococcaceae:
OC Desulfurococcaceae: Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE=99310339; PubMed=10382966;
RA Kawabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka H., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Ogunchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1";
RL DNA Res. 6:83-101(1999).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-tRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000064; BAA81329.1; -
DR HSSP: P00962; 1GTR
DR InterPro: IPR004526; GLX_arch.
DR InterPro: IPR000924; Glu_tRNA-synt_1c.
DR InterPro: IPR001412; tRNA-synt_1.
DR Pfam: PF00749; tRNA-synt_1c; 1.
DR PRINTS: PR00987; TRNASYNTHGLD.
DR TIGRFAMs: TIGR00463; gltx_arch; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 79 89 "HIGH" REGION.
FT SQT 544 AA; 62894 MW; 52B85CDB3EBA984 CRC64;
SQ SEQUENCE 544 AA; 62894 MW; 52B85CDB3EBA984 CRC64;
Query Match 12.3%; Score 69.5; DB 1; Length 544;
Best Local Similarity 35.9%; Pred. No. 5;
Matches 28; Conservative 5; Mismatches 30; Indels 15; Gaps 4;
QY 3 WD--MFQSTHPCAFHNASRAISGPIYSDSVGKHNFDLKLVLDPDGSILRSEYVALP 60
DB 140 WDEYIOSLMEVFYVSARAIRGCAIV-DNCRGCKELLS-----RGEY--CP 186
QY 61 TRDCLFEDPLHNGETMLK 78
DB 187 TRDLPEDNLEFEKME 204
RESULT 3
FLA3_PYRKO STANDARD; PRT; 258 AA.
AC 09V2W9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin B3 precursor.
GN FLA3.
OS Pyrococcus kodakaraensis.
OC Archaea: Euryarchaeota: Thermococci: Thermococcales: Thermococcaceae:
OC Thermococcus.
OX NCBI_TaxID=69014;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOD1.
RX MEDLINE=99413236; PubMed=10483738;
RA Nagahisa K., Ezaki S., Fujiwara S., Imanaka T., Takegi M.;
RT "Sequence and transcriptional studies of five clustered flagellin
RT genes from hyperthermophilic archaeon Pyrococcus kodakaraensis KOD1";
RL FEMS Microbiol. Lett. 178:183-190(1999).
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF FLAGELLA.
CC -1- SIMILARITY: BELONGS TO THE ARCHAEAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB018434; BAA84107.1; -
DR InterPro: IPR002774; Arch_flagellin.
DR Pfam: PF01917; Arch_flagellin; 1.
KW Flagella; Multigene family.
FT PROPEP 1 8 BY SIMILARITY.
FT CHAIN 9 258 FLAGELLIN B3.
SQ SEQUENCE 258 AA; 27443 MW; E0AD08C1B7ABBEAC CRC64;
Query Match 11.9%; Score 67; DB 1; Length 258;
Best Local Similarity 28.3%; Pred. No. 4.1;
Matches 26; Conservative 15; Mismatches 29; Indels 22; Gaps 4;
QY 20 RAISGCPYVSDSVGKHNFDLKLVLDPDGSILRSEYVALPTR 62
DB 77 KPISKLAIVSPNAGSGIDMKKRYVLSDGSTEAVLKYTONSDSDGTLDKQVFAVGMP 136
QY 63 DCLFED----PLHNGETMLKIMNLN-KFTGVI 89
DB 137 DNVEFDDTGTAAVDGDQYITWSELNDKTFGLI 168
RESULT 4
MRZ2_TREPA STANDARD; PRT; 149 AA.
AC 083398;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein mrz2.
GN MRZ2 OR TP0383.
OS Treponema pallidum.
OC Bacteria: Spirochaetales: Spirochaetaceae: Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Winn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod W.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ueberlack T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith R.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete";
RL Science 281:375-388(1998).
CC -1- SIMILARITY: BELONGS TO THE MRZ2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

FT VARIANT 175 175 R -> G (IN SILVER).
 FT VARIANT 373 373 D -> N (IN SILVER).
 FT VARIANT 471 471 F -> S (IN SILVER).
 FT VARIANT 603 626 AAPASGLRARGLENSPLSGOOV -> SSASLRSSRPBW
 FT SEQUENCE 626 AA: 65980 MW: 7AB94ID2E3FB1044 CRC64;
 RKPAPQMTAGLIIKAPMTSMG (IN SILVER).

Query Match 11.4%; Score 64.5; DB 1; Length 626;
 Best Local Similarity 24.8%; Pred. No. 21;
 Matches 29; Conservative 17; Mismatches 42; Indels 29; Gaps 7;

QY 1 PDDMDQSTHPCAFHAASRAISGPIYV-----SDSYGKINFDLKKLVLPDGSIL----- 52
 DB 50 PEWTEVQGSNCMRGGVSLVINDGPTLVGAMASFSIALH-FPGSQK-VLPDQGVIMANN 107
 OY 53 -----RSEYVALPTRDCLFED--PLHNG-----ETMLKIMVL-NKFTGVIG 90
 DB 108 TIINGSGVWGQVRYOPEPDACVFPDGGFCPSGPRKPSFYVWKTKGKYMVLG 164

RESULT 7

TS1L_GIALA STANDARD; PRT: 667 AA.
 AC 003185;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Major surface trophozoite antigen 11 precursor.
 GN TSPL1.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
 OX NCBI_TaxID=5741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ISolate AD-1;
 RX MEDLINE=93241215; PubMed=8479449;
 RA Ey P.L., Khanna K., Manning P.A., Mayrhofer G.
 RT "A gene encoding a 69-kilodalton major surface protein of Giardia
 RL Mol. Biochem. Parasitol. 58:247-258(1993).
 CC -1 SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
 CC PLASMA MEMBRANE.
 CC -1 DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.
 CC -1 SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: M95814; AAA02687.1; -
 CC PIR: A48579; A48579.
 DR HSSP: P02468; IKLO.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR005127; Giardia_VSP.
 DR Pfam: PF03302; VSP; 1.
 DR SMART: SM00181; EGF; 3.
 DR SMART: SM00261; FUF; 5.
 DR Antigen; Repeat; Transmembrane; Signal.
 KW SIGNAL.
 FT CHAIN 18 667 MAJOR SURFACE TROPHOZOITE ANTIGEN 11.
 FT DOMAIN 18 633 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 634 662 POTENTIAL.
 FT DOMAIN 663 667 CYTOSOLASMIC (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 667 AA: 68475 MW: 1DD95727032328BD CRC64;

Query Match

11.4%; Score 64.5; DB 1; Length 667;

Best Local Similarity 30.0%; Pred. No. 22;
 Matches 18; Conservative 7; Mismatches 18; Indels 17; Gaps 3;

OY 45 VLDDG--SILRSRYVLPTRD-----CLFEDPLHNGETMLKIMLNKFGVIAFNC 94
 DB 89 VSPDGVCSVANEYFVPPNADATHDSVSCSEETPIH-----LANKKQYIGVACATC 141

Query Match 11.3%; Score 64; DB 1; Length 105;
 Best Local Similarity 21.9%; Pred. No. 3.2;
 Matches 25; Conservative 13; Mismatches 32; Indels 44; Gaps 4;

GLPE_HAEIN STANDARD; PRT: 105 AA.
 AC P44819;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Thiosulfate sulfotransferase glpe (EC 2.8.1.1).
 GN GLPE OR HI0679.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7342800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RL Science 269:496-512(1995).
 CC -1 FUNCTION: Catalyzes, although with low efficiency, the sulfur
 CC transfer reaction from thiosulfate to cyanide (By similarity).
 CC -1 CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1 SIMILARITY: BELONGS TO THE GLPE FAMILY.
 CC -1 SIMILARITY: CONTAINS 1 RHODANESE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U32750; AAC2338.1; -
 CC HSSP: P09390; IGMX.
 DR TIGR: HI0679; -
 DR InterPro: IPR001763; Rhodanese-like.
 DR Pfam: PF00581; Rhodanese; 1.
 DR SMART: SM00450; RHOD; 1.
 KW Transferase; Glycerol metabolism; Complete proteome.
 FT ACT SITE 63 63 BY SIMILARITY.
 SQ SEQUENCE 105 AA: 12033 MW: A3120E14BDD7245F CRC64;

Query Match

Best Local Similarity 21.9%; Pred. No. 3.2;
 Matches 25; Conservative 13; Mismatches 32; Indels 44; Gaps 4;

OY 3 WDMFGS-----THPCAFAHAAASRAISGPIYVSDSYGKINFDLKKLVLP 47
 DB 12 WEMMOGAGAILVDIRDMRPRAYSHPKGAFHLTNO-----SFLQFEELADF 55
 OY 48 DGSILRSEYVALPTRDCLFEDPLHNGETMLKIMLNKFTGVIGAFNCGGGGCR 101
 DB 56 DSEPIVSCYHGVSSR-----NVATFLVEQGYKNVPSMTGGFD-----GMCR 96

```
RESULT 9
VN3A_ROTIPC STANDARD: PRT: 402 AA.
AC P27586;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Nonstructural RNA-binding protein 34 (NS34) (NCVPA).
GN S6.
OS Porcine rotavirus (group C / strain Cowden).
OC Viruses; dsRNA viruses; Reoviridae; Rotaviruses.
OX NCBI_TaxID=10916;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91361567; PubMed=1653496;
RA Qian Y.A., Jiang B.M., Sait L.J., Kang S.Y., Ojeh C.K., Green K.Y.,
RT "Molecular analysis of the gene 6 from a porcine group C rotavirus
RL that encodes the NS34 equivalent of group A rotaviruses.";
CC Virology 184:752-757(1991).
CC -1- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M69115; AAA47087.1; -.
CC PIR: A41040; MNXRPC.
CC InterPro: IPR001159; DS_RBD.
CC InterPro: IPR002873; Rota_NSP3.
CC DR Pfam: PF00035; dsrm: 1.
CC DR Pfam: PF01665; Rota_NSP3: 1.
CC DR ProSITE: PS50137; DS_RBD: 1.
CC KW Nonstructural protein; RNA-binding.
CC FT DOMAIN 384 400 DBM.
CC SEQUENCE 402 AA; 45125 MW; 8B11F7AC7FB5135C CRC64;

Query Match 11.3%; Score 63.5; DB 1; Length 402;
Best Local Similarity 27.5%; Pred. No. 16;
Matches 28; Conservative 17; Mismatches 40; Indels 17; Gaps 5;

OY 8 STHPCAFAHAAASRAISGPIYVSDVSKHNFD-----LLKKLYLDGSLRSEYYA 58
DB 201 STNSTAAFYVS---TIVGNPIKMYDESGKPLFDVGDYLNPKHIIDKMIENIEPIFKSDYRN 257
OY 59 LPTRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCQGGGMC 100
DB 258 NESPD--FD--VMNERSNLIKIVSIND-CHALICVFKEFNAMWC 294

RESULT 10
BBS2_RAT STANDARD: PRT: 721 AA.
AC 099WH9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bardet-Biedl syndrome 2 protein homolog.
GN BBS2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21181710; PubMed=11285252;
RA Nishimura D.Y., Seardy C.C., Carmi R., Elbedour K., Van Maldergem L.,
```

```
RA Fulton A.B., Lam B.L., Powell B.R., Swiderski R.E., Bugge K.E.,
RA Haider N.B., Kwitek-Black A.E., Ying L., Duhl D.M., Gorman S.M.,
RA Heon E., Iannaccone A., Bonneau D., Biesecker L.G., Jacobson S.G.,
RA Stone E.M., Sheffield V.C.;
RT "Positional cloning of a novel gene on chromosome 16q causing
RT Bardet-Biedl syndrome (BBS2).";
RL Hum. Mol. Genet. 10:865-874(2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF342738; AAK28554.1; -.
CC SEQUENCE 721 AA; 79917 MW; 55935DE19C0509DC CRC64;
SQ
```

```
Query Match 11.3%; Score 63.5; DB 1; Length 721;
Best Local Similarity 24.4%; Pred. No. 32;
Matches 21; Conservative 19; Mismatches 17; Indels 29; Gaps 6;

OY 6 FOSTHPCAFAHAAASRAISGPIYVSDVSKHNFDLLKKLYLPDGSILNSEYYALPTROCL 65
DB 24 YDGTHPCLAA--AATQA---GKVF-----HN-----PHMSQHFSTR---V 57
OY 66 FEDPLHNGETMLKINLNKFTGVICA 91
DB 58 FQSPLESDVSL--NINQVSCIGA 80
```

```
RESULT 11
Y083_NPVOP STANDARD: PRT: 819 AA.
AC 010336;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 91.1 kDa protein (ORF86).
DE Oryzias pseudoscutigata multicapsid polyhedrosis virus (OPMPV).
OS Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohmann G.F.;
RT "The sequence of the Oryzias pseudoscutigata multineocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U75930; AAC59085.1; -.
CC InterPro: IPR002557; Chitin_bind_Pera.
CC Pfam: PF01607; CBM_14: 2.
CC DR SMART: SM00494; ChIBD2; 1.
CC KW Hypothetical protein.
CC SEQUENCE 819 AA; 91067 MW; 108942AF0A8F925A CRC64;
SQ
```

```
Query Match 11.3%; Score 63.5; DB 1; Length 819;
Best Local Similarity 23.4%; Pred. No. 36;
Matches 25; Conservative 10; Mismatches 29; Indels 43; Gaps 5;
```

OY 1 PDWDFQ-----STHCAFAASRAISGSPVDSVSGHNFDLKLKVLPGSILR 53
 Db 261 PDQVFDRAIMTCVTHHC-AFNGA-----GHTITADIG----- 294
 OY 54 SEYVALPTRDCLFEDPLHNGETMLKIWNKFTYGIGAFNCGGSGC 100
 Db 295 -----DAOFFKCLNDREAOI-ITCINVRGADGAGYACSGDARC 331
 RESULT 12
 TERT_YEAST STANDARD. PRT. 884 AA.
 AC Q06163;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Telomerase reverse transcriptase (EC 2.7.7.-) (Telomerase catalytic subunit).
 GN EST2 OR YLR318W OR L8543.12.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota, Fungi, Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RP [1]
 RP SEQUENCE FROM N.A.7.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313267; PubMed=9169871;
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoerge W., Benes V., Bruckner M., Delius H., Dupois E., Dusterhoft A., Ertan K.D., Floeth M., Goffeau A., Hebling U., Heumann K., Heuss-Meitzel D., Hilbert H., Hilger F., Kleine K., Kotter P., Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mostl D., Mueller-Auer S., Neulwich U., Obermaler B., Piravandi E., Pohl T.M., Portetlelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M., Scharte M., Scherens B., Scholler P., Schwager C., Schwart S., Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E., Wedler H., Zimmermann F.K., Zollner A., Hanl J., Honetzel J.D.;
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
 RT Nature 387:87-90(1997).
 RL [2]
 RN CHARACTERIZATION.
 RP MEDLINE=97274210; PubMed=9110970;
 RX Lingner J., Hughes T.R., Shevchenko A., Mann M., Lundblad V., Roca T.R.;
 RA "Reverse transcriptase motifs in the catalytic subunit of telomerase.";
 RT Science 276:561-567(1997).
 RL -1- FUNCTION: TELOMERASE IS A RIBONUCLEOPROTEIN ENZYME ESSENTIAL FOR THE REPLICATION OF CHROMOSOME TERMINI IN MOST EUKARYOTES. IT ELONGATES TELOMERES. IT IS A REVERSE TRANSCRIPTASE THAT ADDS SIMPLE SEQUENCE REPEATS TO CHROMOSOME ENDS BY COPYING A TEMPLATE SEQUENCE WITHIN THE RNA COMPONENT OF THE ENZYME.
 CC -1- SUBCELLULAR LOCATION: Nucleus.
 CC -1- MISCELLANEOUS: DELETION CAUSES TELOMERE SHORTENING AND SENESCENCE.
 CC -1- SIMILARITY: BELONGS TO THE REVERSE TRANSCRIPTASE FAMILY.
 CC TELOMERASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb.ch/announce/isb.ch>).
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL: U20618; AAB64520.1; -;
 DR SGD: S0004310; EST2.
 DR InterPro: IPR000477; RYTse.
 DR InterPro: IPR003545; Telomerase_RT.
 DR Pfam: PF00078; tvt. 1.
 DR PRINTS: PRO1365; TELOMERASERT.
 DR

```

KW Transferrase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
KM DNA-binding.
SQ SEQUENCE      884 AA; 102663 MW; 788334BB49592340 CRC64;

Query Match          11.2% Score 63; DB 1; Length 884;
Best Local Similarity 29.5%; Pred. No. 45;
Matches    23; Conservative   15; Mismatches    30; Indels     10; Gaps       4;

QY      26 PIYSDSYGKINFDLTKL--VLPGDSL---PSEXYALPTROC--LFEDPL--HNGET 75
Db      499 PTQADRIKERORILKKFNVLPELYFMKDKVKSCTSIDPRMECRILKALKLNENGF 558
        |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY      76 MKIWNLEKFTGVIGAFN 93
Db      559 VRSQYFNTMTGTGLKFN 576
        :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 13
NIA_NEUCR STANDARD; PRT; 982 AA.
AC POB619;
DT 01-AUG-1988 (Rel. 08, Created)
DF 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Nitrate reductase [NADPH] (EC 1.6.6.3) (NR).
GN NIT-3.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OX Sordariales; Sordariaceae; Neurospora.
CX NCBI_Taxid=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=91287699; PubMed=1829499;
RA Okamoto P.M., Fu Y.-H., Marzluf G.A.;
RT "Nit-3, the structural gene of nitrate reductase in Neurospora
RT crassa: nucleotide sequence and regulation of mRNA synthesis and
RT turnover.";
RL Mol. Gen. Genet. 227:213-223(1991).
RN [2]
RP PRELIMINARY PARTIAL SEQUENCE AROUND HIS-65Z.
RA Le K.H.D., Lederer F.;
RT "On the presence of a heme-binding domain homologous to cytochrome b5
RT in Neurospora crassa assimilatory nitrate reductase.";
RL EMBO J. 2:1909-1914(1983).
RN [3]
RP MUTANTS.
RC STRAIN=74-OR23-1A;
RX MEDLINE=93241176; PubMed=8479443;
RA Okamoto P.M., Garrett R.H., Marzluf G.A.;
RT "Molecular characterization of conventional and new repeat-induced
RT mutants of nit-3, the structural gene that encodes nitrate reductase
RT in Neurospora crassa.";
RL Mol. Gen. Genet. 238:81-90(1993).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=93360901; PubMed=8355655;
RA Okamoto P.M., Marzluf G.A.;
RT "Nitrate reductase of Neurospora crassa: the functional role of
RT individual amino acids in the heme domain as examined by
RT site-directed mutagenesis.";
RL Mol. Gen. Genet. 240:221-230(1993).
CC -1- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -1- CATALYTIC ACTIVITY: NADPH + nitrate = NADP(+) + nitrite + H(2)O.
CC -1- COFACTOR: EACH SUBUNIT CONTAINS 1 EQUIVALENT OF FeD, HEHE IRON,
CC AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME GROUP IS
CC CALLED CYTOCHROME B-557.
CC -1- PATHWAY: NITRATE ASSIMILATORY PATHWAY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- INDUCTION: ITS EXPRESSION IS HIGHLY REGULATED AND RESPONDS RAPIDLY
CC TO NITRATE INDUCTION AND TO NITROGEN REPRESSION.
CC -1- SIMILARITY: TO EUKARYOTIC MOLYBDOPROTEIN OXIDOREDUCTASES IN THE

```



```
CC N-TERMINAL DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
CC -1- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE C-TERMINAL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X61303; CAA43600.1; -.
CC DR PIR: S16292; S16292.
CC DR PIR: S34796; S34796.
CC DR PIR: S37298; S37298.
CC DR HSSP: P00171; IFO3.
CC DR InterPro: IPR001199; Cytochrome b5 reductase.
CC DR InterPro: IPR001834; Cytochrome b5 reductase.
CC DR InterPro: IPR000572; Enzyme oxidoreductase.
CC DR InterPro: IPR005066; Mo-co-dimer.
CC DR InterPro: IPR001433; Oxidoreductase/FAD/NAD(P).
CC DR Pfam: PF00173; heme_1; 1.
CC DR Pfam: PF00174; oxidored_molyb; 1.
CC DR Pfam: PF00175; NAD_binding_1.
CC DR Pfam: PF00970; FAD_binding_6; 1.
CC DR Pfam: PF03404; Mo-co_dimer; 1.
CC DR PRINTS: PR00406; CYTB5RDYSE.
CC DR PRINTS: PR00363; CYTOCHROMB5.
CC DR PRINTS: PR00407; EUMOPPERIN.
CC DR ProDom: PD00612; Cytochrome b5; 1.
CC DR PROSITE: PS00191; CYTOCHROME_B5_1; 1.
CC DR PROSITE: PS00255; CYTOCHROME_B5_2; 1.
CC DR PROSITE: PS00559; MOLYBDOPTERIN_EUK; 1.
CC KW Oxidoreductase; Flavoprotein; FAD; NADP; Heme; Molybdenum; Nitrate assimilation.
KW METAL.
FT 240 240
FT METAL 295 295
FT DISULFID 499 499
FT DOMAIN 621 691
FT BINDING 652 652
FT BINDING 675 675
FT DOMAIN 718 982
FT NP_BIND 952 961
FT MUTAGEN 652 652
FT MUTAGEN 675 675
SQ SEQUENCE 982 AA: 108432 MW: 87836C031B19687F CRC64;

Query Match 11.2%; Score 63; DB 1; Length 982;
Best Local Similarity 24.6%; Pred. No. 51;
Matches 31; Conservative 11; Mismatches 38; Indels 46; Gaps 6;

OY 1 PDMDMEQ--STRPCAFHSAISRAISGCPRIYSDSVCKHNFDLLKTLVLPDGSITLSE---- 55
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 781
DB 736 PDKTKIFHSLHPAOSI-----GLP-----VGQHLMMLRPPAPAKTESIIIRAYPI 781
OY 56 ----YYALPTDCLEDFEDPLHNGE-----TWLKTIWNLNKFPGVIGAEFC 94
DB 782 SDCTLERGTLRVLVKVIYYIASPED-----IKGGOMTALDALGAKVAEEFGPKGFYV 835
OY 95 OCGGWC 100
DB ||| |
DB 836 OGRGVC 841

RESULT 14
ID Y872_RICPR STANDARD; PRT; 165 AA.
AC O92C94;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE Hypothetical protein RP872.
GN R8972.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxId=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sieberitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naestlund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria."
RL Nature 396:133-140(1998).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch)
CC -----
CC DR EMBL; AJ235273; CAAI5296.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 165 AA; 19411 MW; 305290A3A14024C4 CRC64;

Query Match      11.1%; Score 62.5; DB 1; Length 165;
Best Local Similarity 31.5%; Pred. NO. 7.8;
Matches 23; Conservative 12; Mismatches 31; Indels 7; Gaps 4;

QY 26 PIVVSDSVGKHHFDLLKKLVLPD--GSILR-SEYVALPTRODFEDPDLHN-GETMLKIWN 81
Db 83 PIYYIDELQNPNMNAVLYIISTPDIAKLQWKEVIKYVKRRIRIIITDLENLKELIKI-- 140
QY 82 LNKFTVGIGAFNC 94
Db 141 NKFTEGENKIDIC 152

RESULT 15
API2_SOLUTION STANDARD: PRT: 220 AA.
AC O43646;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ASPARTIC PROTEASE INHIBITOR 2 precursor (Cathepsin D inhibitor)
DE (Cathinh) (Cathln).
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxId=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. AM 80.5793; TISSUE=leaf;
RX MEDLINE=95036033; PubMed=7948907;
RA Hebers K., Prat S., Willmitzer L.;
RT "Cloning and characterization of a cathepsin D inhibitor gene from
RL Solanum tuberosum L.;"
RL Plant Mol. Biol. 26:73-83(1994).
CC -1- FUNCTION: INHIBITOR OF CATHEPSIN D (ASPARTIC PROTEASE). MAY ALSO
CC INHIBIT TRYPSIN AND CHYMOTRYPSIN (SERINE PROTEASES). PROTECTS THE
CC PLANT BY INHIBITING PROTEASES OF INVADING ORGANISMS.
CC -1- SUBCELLULAR LOCATION: Vacuolar (By similarity).
CC -1- TISSUE SPECIFICITY: TUBERS.
CC -1- INDUCTION: NOT INDUCED BY ABSICISC ACID, JASMONIC ACID AND
CC WOUNDING.
CC -1- SIMILARITY: BELONGS TO THE LEGUMINOUS KUNITZ-TYPE INHIBITOR
CC FAMILY.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74985; CAA52919.1; -.
CC InterPro: IPR002160; Kunitz_legume.
CC Pfam: PF00197; Kunitz_legume; 1.
CC PRINTS; PR00291; KUNITZINHBTR.
CC PRODOM; PD000891; Kunitz_legume; 1.
CC SMART; SM00452; SPI; 1.
CC PROSITE; PS00283; SOYBEAN_KUNITZ; 1.
CC Aspartic protease inhibitor; Serine protease inhibitor; Signal;
CC Multigene family.
CC SIGNAL 1 23 BY SIMILARITY.
CC PROPEP 24 32 BY SIMILARITY.
CC CHAIN 33 220 ASPARTIC PROTEASE INHIBITOR 2.
CC SITE 26 31 VACUOLAR TARGETING SIGNAL (BY
CC SIMILARITY).
CC ACT_SITE 99 100 REACTIVE BOND (FOR TRYPSIN) (BY
CC SIMILARITY).
CC ACT_SITE 144 145 REACTIVE BOND (FOR CHYMOTRYPSIN) (BY
CC SIMILARITY).
CC DISULFID 80 125 BY SIMILARITY.
CC DISULFID 174 185 BY SIMILARITY.
CC CARBOHYD 51 51 N-LINKED (GLCNAC... ) (POTENTIAL).
CC SEQUENCE 220 AA: 24199 MW: 33D6E866EA0C5DD5 CRC64;

Query Match 11.1%; Score 62.5; DB 1; Length 220;
Best Local Similarity 25.0%; Pred. No. 11;
Matches 25; Conservative 15; Mismatches 27; Indels 33; Gaps 5;

QY 10 HPCAFHNAS--RAISGPIYVS-----DSYGHNFDLKLVLPDGSILRSEY 57
   : : : : | | : | : : : : | : : : : | : : : :
Db 49 NPNSSYRIISIGRGALGGVYLKSPNSDGPDPGVFRYNSD-----VGPSGTFVR----- 99

QY 58 ALPTRDCLFEDPLHN-----GETMLKIVNLNKF 85
   : : : : | | | | | : : : : :
Db 100 FTPLSGIFEDQLNINQENIATVKLCVSTYTKVGNLNAV 139
```

Search completed: March 27, 2003, 10:04:21
Job time : 19.628 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:40 ; Search time 24.6341 Seconds
(without alignments)
120.634 Million cell updates/sec

Title: US-09-675-208-5_COPY_510_610

Perfect score: 564

Sequence: 1 PDMDMFQSTHPCAFHAAASRAISGPIIYSDSVGKINFLKLVLPDGSILRSEYALP 101

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	564	100.0	784	US-08-846-234-5	Sequence 5, Appl
2	85	15.1	174	US-09-153-586-23	Sequence 23, Appl
3	63	11.2	884	US-08-851-843A-55	Sequence 55, Appl
4	63	11.2	884	US-08-974-549A-222	Sequence 222, App
5	63	11.2	884	US-08-854-050-55	Sequence 55, Appl
6	63	11.2	884	US-09-330-323-55	Sequence 55, Appl
7	61	10.8	3729	US-08-804-227C-4	Sequence 4, Appl
8	60.5	10.7	430	PCT-US93-04392-3	Sequence 3, Appl
9	60	10.6	210	US-08-948-616-5	Sequence 5, Appl
10	60	10.6	210	US-09-193-510-5	Sequence 5, Appl
11	60	10.6	210	US-09-368-402-5	Sequence 5, Appl
12	60	10.6	210	US-09-325-932A-179	Sequence 179, App
13	60	10.6	890	US-09-342-648-10	Sequence 10, Appl
14	59.5	10.5	430	PCT-US93-04392-6	Sequence 6, Appl
15	59.5	10.5	430	PCT-US93-04392-9	Sequence 9, Appl
16	59	10.5	390	US-09-586-719-2	Sequence 2, Appl
17	58.5	10.4	430	PCT-US93-04392-12	Sequence 12, Appl
18	58.5	10.4	430	PCT-US93-04392-15	Sequence 15, Appl
19	58.5	10.4	536	US-08-426-509A-12	Sequence 12, Appl
20	58.5	10.4	536	PCT-US95-05008-12	Sequence 12, Appl
21	58	10.3	210	US-08-948-616-10	Sequence 10, Appl
22	58	10.3	210	US-08-991-946A-6	Sequence 6, Appl
23	58	10.3	210	US-09-193-510-10	Sequence 10, Appl
24	58	10.3	210	US-09-368-402-10	Sequence 10, Appl
25	58	10.3	226	US-08-440-517A-6	Sequence 6, Appl
26	58	10.3	226	US-09-092-160-6	Sequence 6, Appl
27	58	10.3	228	US-08-766-982-11	Sequence 11, Appl

28	58	10.3	228	US-08-944-483-55	Sequence 55, Appl
29	58	10.3	228	US-09-296-219-11	Sequence 11, Appl
30	58	10.3	711	US-08-184-012C-8	Sequence 8, Appl
31	58	10.3	711	US-08-334-177-2	Sequence 2, Appl
32	58	10.3	711	US-08-666-082B-1	Sequence 1, Appl
33	58	10.3	711	US-08-766-982-2	Sequence 2, Appl
34	58	10.3	711	US-08-948-616-10	Sequence 10, Appl
35	58	10.3	711	US-09-296-219-2	Sequence 2, Appl
36	58	10.3	1584	PCT-US95-13830-2	Sequence 2, Appl
37	57.5	10.2	537	US-09-251-645-6	Sequence 6, Appl
38	57.5	10.2	537	US-08-426-509A-11	Sequence 11, Appl
39	57.5	10.2	954	PCT-US95-05008-11	Sequence 11, Appl
40	57	10.1	256	US-09-057-969-2	Sequence 2, Appl
41	57	10.1	256	US-08-484-905-114	Sequence 114, App
42	57	10.1	256	US-08-481-985B-114	Sequence 114, App
43	57	10.1	256	US-08-370-476-114	Sequence 114, App
44	57	10.1	548	US-09-255-828-24	Sequence 24, Appl
45	57	10.1	858	US-09-255-829-22	Sequence 22, Appl
			858	US-09-255-829-29	Sequence 29, Appl

ALIGNMENTS

```
RESULT 1
US-08-846-234-5
: Sequence 5, Application US/08846234
: Patent No. 6166292
: GENERAL INFORMATION:
: APPLICANT: OSUMI Chieko
: APPLICANT: NOZAKI Jinshi
: APPLICANT: KIDA Takao
: TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
: TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
: STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
: CITY: ARLINGTON
: STATE: VIRGINIA
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/846,234
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: NORMAN F. OBLON
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)-413-3000
: TELEFAX: (703)-413-2220
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 784 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-846-234-5

Query Match      100.0%      Score 564;      DB 4;      Length 784;
Best Local Similarity 100.0%;      Pred. No. 1.2e+64;
Matches 101;      Conservative 0;      Mismatches 0;      Indels 0;      Gaps 0;

OY      1 PDMDMFQSTHPCAFHAAASRAISGPIIYSDSVGKINFLKLVLPDGSILRSEYALP 60
      |||
Db      510 PDMDMFQSTHPCAFHAAASRAISGPIIYSDSVGKINFLKLVLPDGSILRSEYALP 569
OY      61 TRDCLFEDPLHNGETMLKLTWLNKFTGVGAENCGGCMCR 101
      |||
```



```

1      NUMBER OF SEQUENCES: 225
2      CORRESPONDENCE ADDRESSES:
3      ADDRESSEE: Townsend and Townsend and Crew LLP
4      STREET: Two Embarcadero Center, 8th Floor
5      CITY: San Francisco
6      STATE: California
7      COUNTRY: United States of America
8      ZIP: 94111
9
10     COMPUTER READABLE FORM:
11     MEDIUM TYPE: Floppy disk
12     COMPUTER: IBM PC compatible
13     OPERATING SYSTEM: PC-DOS/MS-DOS
14     SOFTWARE: PatentIn Release #1.0, Version #1.30
15     CURRENT APPLICATION DATA:
16     APPLICATION NUMBER: US/09/430,323
17     FILING DATE: 29-Oct-1999
18     CLASSIFICATION: <unknown>
19
20     PRIOR APPLICATION DATA:
21     APPLICATION NUMBER: US 08/854,050
22     FILING DATE: 09-MAY-1997
23     APPLICATION NUMBER: US 08/851,843
24     FILING DATE: 06-MAY-1997
25     APPLICATION NUMBER: US 08/846,017
26     FILING DATE: 25-APR-1997
27     APPLICATION NUMBER: US 08/844,419
28     FILING DATE: 18-APR-1997
29     APPLICATION NUMBER: US 08/724,643
30     FILING DATE: 01-OCT-1996
31     ATTORNEY/AGENT INFORMATION:
32     NAME: Apple, Randolph T.
33     REGISTRATION NUMBER: 36,429
34     REFERENCE/DOCKET NUMBER: 015389-00293005
35     TELECOMMUNICATION INFORMATION:
36     TELEPHONE: (415) 576-0200
37     TELEFAX: (415) 576-0300
38     INFORMATION FOR SEQ ID NO: 55:
39     SEQUENCE CHARACTERISTICS:
40     LENGTH: 884 amino acids
41     TYPE: amino acid
42     STRANDEDNESS: not relevant
43     TOPOLOGY: not relevant
44     MOLECULE TYPE: peptide
45     SEQUENCE DESCRIPTION: SEQ ID NO: 55:
46
47     US-09-430-323-55
48
49     Query Match      11.2%: Score 63; DB 4; Length 884;
50     Best Local Similarity 29.5%: Pred. NO. 24;
51     Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4.
52
53     Oy      26  PIYVDSVQKHNFDLKLK-LLPGSIL----RSEYVALPTRDC--LPEDDL--HNGFT 75
54             |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
55     Db      499  PQIADRIKEFQRLRLKFNVLPELYFAKFDVKSCIDSIPIAMECWRIILKDALKNNGFF 558
56             :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
57     Oy      76  MKIWNLKFTGVIGAFN 93
58             :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
59     Db      559  VRSQYFENTMGVLKLFN 576
60
61     RESULT 7
62     US-08-804-227C-4
63     Sequence 4, Application US/08804227C
64     Patent No. 5876991
65     GENERAL INFORMATION:
66     APPLICANT: Dehoff, Bradley S.
67     APPLICANT: Kuhstoss, Stuart A.
68     APPLICANT: Kosteck, Paul R., Jr.
69     APPLICANT: Sutton, Kimberly L.
70     TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
71     NUMBER OF SEQUENCES: 15
72     CORRESPONDENCE ADDRESSES:
73     ADDRESSEE: THOMAS G. PLANT 1501
74     STREET: LILLY CORPORATE CENTER
75     CITY: INDIANAPOLIS

```

```

1 STATE: IN
2 COUNTRY: USA
3 ZIP: 46285
4
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM Compatible
8 OPERATING SYSTEM: MS-DOS
9 SOFTWARE: ASCII(DOS) Text only
10
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/804,227C
13 FILING DATE: February 21, 1997
14
15 CLASSIFICATION: 435
16
17 ATTORNEY/AGENT INFORMATION:
18 NAME: Plant, Thomas, G.
19 REGISTRATION NUMBER: 35,784
20 REFERENCE/DOCKET NUMBER: X-8231
21
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: 317-276-2459
24
25 INFORMATION FOR SEQ ID NO: 4:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 3729 amino acids
28 TYPE: amino acid
29 TOPOLOGY: unknown
30
31 MOLECULE TYPE: protein
32
33 US-08-804-227C--4
34
35 Query Match 10.8%; Score 61; DB 2; Length 3729;
36 Best Local Similarity 30.9%; Pred. No. 3.2e+02;
37 Matches 17; Conservative 13; Mismatches 21; Indels 4; Gaps 3
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

ORIGINAL SOURCE:
ORGANISM: Zopfiella laclpes
STRAIN: 511
INDIVIDUAL ISOLATE: ATCC #26183
IMMEDIATE SOURCE:
CLONE: ZL511 1-2
PCR-US93-04392-3

Query Match 10.7%; Score 60.5; DB 5; Length 430;
Best Local Similarity 25.5%; Pred. No. 18;
Matches 25; Conservative 15; Mismatches 39; Indels 19; Gaps 4;

OY 10 HPCAFHNASRAISGC--PIYSD-----SVGKHNFDLKLVLPDGSILRSEYVA 58
DB 248 HPDDPLATGQYVLAVGGNNMLVADGDFGCHGFTTGENTYLVKLSLLANDGKLLSPMVN 307
OY 59 LPRDCLFEDPLNGETMLKTMNKFTGYTGAFNCOG 96
DB 308 L-----MFEDHLTG---AKKGHDALNGVGSFFAVG 337

RESULT 9

US-08-948-616-5
Sequence 5, Application US/08948616
Patent No. 5840539
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Puryl
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,616
FILING DATE: Herewith
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 210 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KERANOT02
CLONE: 2620104
US-08-948-616-5

Query Match 10.6%; Score 60; DB 2; Length 210;
Best Local Similarity 33.3%; Pred. No. 7.8;
Matches 11; Conservative 10; Mismatches 10; Indels 2; Gaps 1;

OY 22 ISGGPI--YVSDSVGKHNFDLKLVLPDGSIL 52
DB 60 VKGKPVHLHWIDTAGQDDYDRLRLFPDASVL 92

RESULT 10

US-09-193-510-5
Sequence 5, Application US/09193510
Patent No. 5981226
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Puryl
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/193,510
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/948,616
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0409 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 210 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: KERANOT02
CLONE: 2620104
US-09-193-510-5

Query Match 10.6%; Score 60; DB 2; Length 210;
Best Local Similarity 33.3%; Pred. No. 7.8;
Matches 11; Conservative 10; Mismatches 10; Indels 2; Gaps 1;

OY 22 ISGGPI--YVSDSVGKHNFDLKLVLPDGSIL 52
DB 60 VKGKPVHLHWIDTAGQDDYDRLRLFPDASVL 92

RESULT 11

US-09-368-402-5
Sequence 5, Application US/09368402
Patent No. 6200790
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Shah, Puryl
APPLICANT: Corley, Neil C.

```

1 TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
2
3 NUMBER OF SEQUENCES: 11
4
5 CORRESPONDENCE ADDRESS:
6
7 ADDRESSEE: Incyte Pharmaceuticals, Inc.
8
9 STREET: 3174 Porter Drive
10
11 CITY: Palo Alto
12
13 STATE: CA
14
15 COUNTRY: USA
16
17 ZIP: 94304
18
19 COMPUTER READABLE FORM:
20
21 MEDIUM TYPE: Diskette
22
23 COMPUTER: IBM Compatible
24
25 OPERATING SYSTEM: DOS
26
27 SOFTWARE: FASTSEQ for Windows Version 2.0
28
29 CURRENT APPLICATION DATA:
30
31 APPLICATION NUMBER: US/09/368,402
32
33 FILING DATE:
34
35 CLASSIFICATION:
36
37 PRIOR APPLICATION DATA:
38
39 APPLICATION NUMBER: 09/193,510
40
41 FILING DATE:
42
43 ATTORNEY/AGENT INFORMATION:
44
45 NAME: Billings, Lucy J.
46
47 REGISTRATION NUMBER: 36,749
48
49 REFERENCE/DOCKET NUMBER: PF-0409 US
50
51 TELECOMMUNICATION INFORMATION:
52
53 TELEPHONE: 650-855-0555
54
55 TELEFAX: 650-845-4166
56
57 TELEEX:
58
59 INFORMATION FOR SEQ ID NO: 5:
60
61 SEQUENCE CHARACTERISTICS:
62
63 LENGTH: 210 amino acids
64
65 TYPE: amino acid
66
67 STRANDEDNESS: single
68
69 TOPOLOGY: linear
70
71 IMMEDIATE SOURCE:
72
73 LIBRARY: KERANOT02
74
75 CLONE: 2620104
76
77 US-09-368-402-5

```

Query Match	10.68;	Score 60;	DB 4;	Length 210;
Best Local Similarity	33.38;	Pred. No. 7.8;		
Matches 11;	Conservative 10;	Mismatches 10;	Indels 2;	Gaps 1

```
QY      22  ISGGPI--YVSDSVGKHNFDLLKLVLPDGSIL 52
          : | | : : | : : | | | : |
Db      60  VKGKPVHLHINVTAGDDYDRLRLPLFYPDASVL 92
```

```

RESULT 12
US-09-325-932A-179
: Sequence 179, Application US/09325932A
: Patent No. 6451604
: GENERAL INFORMATION:
: APPLICANT: Flinn, Barry
: APPLICANT: Lasham, Annette
: TITLE OF INVENTION: Compositions affecting programmed cell
: TITLE OF INVENTION: death and their use in the modification of forestry plant develop
: FILE REFERENCE: 1022
: CURRENT APPLICATION NUMBER: US/09/325,932A
: CURRENT FILING DATE: 1999-06-04
: NUMBER OF SEQ ID NOS: 206
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 179
: LENGTH: 608
: TYPE: PRF
: ORGANISM: Pinus radiata
: US-09-325-932A-179

```

Query Match	10.6%;	Score 60;	DB 4;	Length 608;
Best Local Similarity	21.4%;	Pred. NO. 34;		
Matches 25;	Conservative 18;	Mismatches 48;	Indels 26;	Gaps 5;

```

Oy  3  MDMSOSTHPCAAFHAA-----SKAISGGPPLVSDVGKHNFDLK-KLVLPP 48
Db  427  WHAAGREPRPVVGEKVEKFLMACNGCLPLSLKVLGALLHGDMDKLMAQAQRKSKYLPE 4866
Oy  49  G-SLISEYIALTPRD-----CLFEDPLHNGETMLKTNWLNKNTGYVIGAFNCQ 95
Db  487  DIRSLRLRYSLALKEEKQIFLDIACFFIG--KKRDSAIRWDSQSNNEGGLIGMKLE 541

```

RESULT 13
 US-09-342-648-10
 Sequence 10, Application US/09342648
 Patent No. 6248584
 GENERAL INFORMATION:
 APPLICANT: Cahoon, Rebecca E.
 APPLICANT: Odell, Joan
 APPLICANT: Rafalski, Antoni
 TITLE OF INVENTION: Transcription Coactivators
 FILE REFERENCE: BR-1169-B
 CURRENT FILING DATE: US/09/342,648
 CURRENT FILING DATE: 1999-06-29
 EARLIER APPLICATION NUMBER: 60/092,659
 EARLIER FILING DATE: July 13, 1998
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 10
 LENGTH: 890
 TYPE: prt
 ORGANISM: Ajellomyces capsulatus
 US-09-342-648-10

Query Match	10.6%	Score 60;	DB 4;	Length 890;
Best Local Similarity	25.0%;	Pred. No. 59;		
Matches	19;	Conservative	9;	Mismatches 22;
				Indels 26;
				Gaps 3

OY 48 DGSLSSEVVALPTROCE-----PLHNETML-KITWLNFE 85
|||: |||: | : - : | |||: | :
Db 39 DGSLSYSVALAQFANQSRFLRELLGVKVFQYDVTPTRNGNTMIVLHNQEN 98
OY 86 TGYIGAFCOGGSMCR 101
.: .: ||:
Db 99 LAEL---CVAEGLVK 110

RESULT 14
PCT-US93-04392-6
; Sequence 6, Application PC/TUS9304392

APPLICANT: Beck, Steven R., Cain, Robert O., Chan, Hardy W
APPLICANT: Freedman, Richard, Helfner, Donald L.
APPLICANT: Phelps, Trish, Roberts, Christopher R.
APPLICANT: Salazar, Felix H., Snyder, Roger C.
TITLE OF INVENTION: Enzymatic Process for Production of
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET:
CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04392
FILING DATE: 19930514
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/883,658
FILING DATE: MAY 15, 1992

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: AMINO ACID
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: cDNA
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Zopfiella latipes
STRAIN: 780
INDIVIDUAL ISOLATE: ATCC #44575
IMMEDIATE SOURCE:
CLONE: 21780-3A
PCT-US93-04392-6

Query Match 10.5%; Score 59.5; DB 5; Length 430;
Best Local Similarity 25.5%; Pred. No. 25;
Matches 24; Conservative 15; Mismatches 36; Indels 19; Gaps 4;

Qy 10 HPCAFHNASRAISGG--PIYSD-----SVGKHNFDLKKLYLPDGSILRSEYYA 58
Db 248 HPDDPLATGKQVLAGGNNLVADGDFGCHGMFTTGENTLKVLSLANDGKLSPEMYN 307

Qy 59 LPTRDCLFEDPLANGETMLKINLNKFTGVIGAF 92
Db 308 L-----MFEDHLTGG---AKKGHEDALNCPVGSF 333

RESULT 15
PCT-US93-04392-9
Sequence 9, Application PC/TUS9304392
GENERAL INFORMATION:
APPLICANT: Beck, Steven R., Cain, Robert O., Chan, Hardy W.
APPLICANT: Freedman, Richard, Heefner, Donald L.
APPLICANT: Phelephs, Trish, Roberts, Christopher R.
APPLICANT: Salazar, Felix H., Snyder, Roger C.
TITLE OF INVENTION: Enzymatic Process for Production of
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET:
CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04392
FILING DATE: 19930514
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/883,658
FILING DATE: MAY 15, 1992
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: AMINO ACID
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: cDNA
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Zopfiella latipes
STRAIN: 780
INDIVIDUAL ISOLATE: ATCC #44575
IMMEDIATE SOURCE:

LIBRARY: cDNA in lambda-gtl1
CLONE: 21780-m10
PCT-US93-04392-9

Query Match 10.5%; Score 59.5; DB 5; Length 430;
Best Local Similarity 25.5%; Pred. No. 25;
Matches 24; Conservative 15; Mismatches 36; Indels 19; Gaps 4;

Qy 10 HPCAFHNASRAISGG--PIYSD-----SVGKHNFDLKKLYLPDGSILRSEYYA 58
Db 248 HPDDPLATGKQVLAGGNNLVADGDFGCHGMFTTGENTLKVLSLANDGKLSPEMYN 307

Qy 59 LPTRDCLFEDPLANGETMLKINLNKFTGVIGAF 92
Db 308 L-----MFEDHLTGG---AKKGHEDALNCPVGSF 333

Search completed: March 27, 2003, 10:07:36
Job time : 25.6341 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 71.439 Seconds
(without alignments)
188.389 Million cell updates/sec

Title: US-09-675-208-5_COPY_510_610

Perfect score: 564
Sequence: 1 PDMMFQSTHPCAFHAAASR.....LNKFTGVIGAFNCQGGWCR 101

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A.Geneseq_101002:*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	564	100.0	784	19	AAW53570
2	564	100.0	784	20	AAV17417
3	515	91.3	587	19	AAW57888
4	515	91.3	587	20	AAV30144
5	511	90.6	783	23	ABB93664
6	508	90.1	572	20	AAV32075
7	506	89.7	777	20	AAV32074
8	505	89.5	799	19	AAW57886
9	505	89.5	799	20	AAV30142
10	498	88.3	780	22	AAW98659

11	497	88.1	781	19	AAW57887
12	497	88.1	781	20	AAV30143
13	497	88.1	781	22	AAW49400
14	494	87.6	758	21	AAV70978
15	476	84.4	783	20	AAV32073
16	444	78.7	265	20	AAV32072
17	426	75.5	271	19	AAW57889
18	426	75.5	271	20	AAV30145
19	404	71.6	751	21	AAV70980
20	403	71.5	756	21	AAV70975
21	400	70.9	763	21	AAV70977
22	394	69.9	508	21	AAV70974
23	387	68.6	750	20	AAV17418
24	387	68.6	770	21	AAV70976
25	384	68.1	841	21	AAV70981
26	371	65.8	807	23	ABB92806
27	205	36.3	229	21	AAW28628
28	85	15.1	174	20	AAV06803
29	79	14.0	15	20	AAV17421
30	79	14.0	193	21	AAW28629
31	71	12.6	638	22	ABB64380
32	69.5	12.3	412	18	AAW20443
33	69	12.2	426	22	AAW90870
34	69	12.2	615	22	AAU28187
35	69	12.2	615	22	AAW93480
36	68.5	12.1	183	22	AAU33604
37	68.5	12.1	466	22	ABG09425
38	68.5	12.1	466	22	ABG27212
39	67	11.9	1778	22	AAW79480
40	67	11.9	2099	22	AAW78496
41	65.5	11.6	336	23	AAW92001
42	64	11.3	359	23	ABB93733
43	63.5	11.3	207	18	AAW10493
44	63.5	11.3	423	23	ABP27573
45	63.5	11.3	625	20	AAV31978

ALIGNMENTS

RESULT 1		
ID	AAW53570	standard; Protein: 784 AA.
XX		
AC	AAW53570;	
DT	06-JUL-1998	(first entry)
XX		
DE	Cucurbitur raffinose synthase.	
XX		
KW	Cucurbitur; raffinose synthase; sucrose; galactinol.	
XX		
OS	Cucumis sativus.	
XX		
PN	JP10084973-A.	
XX		
PD	07-APR-1998.	
XX		
PF	28-APR-1997;	97JP-0111124.
XX		
PR	26-JUL-1996;	96JP-0198079.
PR	26-APR-1996;	96JP-0107682.
XX		
PA	(AJIN) AJINOMOTO KK.	
XX		
DR	WPI: 1998-264858/24.	
DR	N-PSDB: AAV22250.	
XX		
PT	Raffinose synthase gene - useful for preparation of raffinose in	
PT	transformed plant	
XX		
PS	Claim 3; Pages 17-20; 26pp; Japanese.	
XX		

Soybean raffinose
Amino acid sequenc
Soybean raffinose
Soybean raffinose
Soybean raffinose
Sugarcane raffinose
Soybean raffinose
Corn raffinose syn
Amino acid sequenc
Wheat raffinose sy
Corn raffinose syn
Rice raffinose syn
Corn raffinose syn
Soybean raffinose
Rice raffinose syn
Wheat raffinose sy
Herbicideally activ
Arabidopsis thalia
Peptide Seq ID No:
Soybean raffinose
Arabidopsis thalia
Drosophila melanog
H. pylori cytoplas
C glutamicum prote
Novel human secret
Human protein sequ
Pseudomonas aerugi
Novel human diagno
Novel human diagno
Human protein SEQ
Human protein SEQ
Herbicideally activ
Soluble fused MHC
Streptococcus poly
Mouse melanoma ant

CC The present sequence is cucumber raffinose synthase, which
CC forms raffinose from sucrose and galactinol, has an optimum pH of
CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
CC Iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA;

Query Match 100.0%; Score 564; DB 19; Length 784;
Best Local Similarity 100.0%; Pred. No. 3.3e-63;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHNFDLKLKLVLPDGSILRSEYALP 60

DB 510 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHNFDLKLKLVLPDGSILRSEYALP 569

OY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 101

DB 570 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 610

RESULT 2

AAV17417

ID AAV17417 standard; Protein; 784 AA.

AC AAV17417;

DT 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase.

KW Raffinose synthase; sucrose; galactinol.

OS Cucumis sativus.

PN JP11123080-A.

PD 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

PA (AJIN) AJINOMOTO KK.

DR WPI: 1999-340516/29.

DR N-PSDB; AAV61238.

PT New raffinose synthase gene - for production of raffinose from
PT sucrose and galactinol

PS Claim 2: Page 25-27; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from cucumber.

XX Sequence 784 AA;

Query Match 100.0%; Score 564; DB 20; Length 784;

Best Local Similarity 100.0%; Pred. No. 3.3e-63;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHNFDLKLKLVLPDGSILRSEYALP 60

DB 510 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHNFDLKLKLVLPDGSILRSEYALP 569

OY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 101

DB 570 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 610

RESULT 3

AAW57888

ID AAW57888 standard; Protein; 587 AA.

AC AAW57888;

DT 23-SEP-1998 (first entry)

DE Japanese artichoke raffinose synthetase.

KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; Japanese artichoke.

OS Stachys sieboldii.

PN EP849359-A2.

PD 24-JUN-1998.

PF 18-DEC-1997; 97EP-0122417.

PR 18-DEC-1996; 96JP-0338673.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI Oeda K, Mantanabe E;

DR WPI: 1998-324670/29.

DR N-PSDB; AAV40802.

PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora

PS Claim 1: Page 36-38; 44pp; English.

CC This sequence is the Japanese artichoke raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.

XX Sequence 587 AA;

Query Match 91.3%; Score 515; DB 19; Length 587;

Best Local Similarity 89.1%; Pred. No. 4.4e-57;
Matches 90; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHNFDLKLKLVLPDGSILRSEYALP 60

DB 401 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHNFDLKLKLVLPDGSILRSEYALP 460

OY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 101

DB 461 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 501

RESULT 4

AAV30144

ID AAV30144 standard; Protein; 587 AA.

AC AAV30144;

DT 26-OCT-1999 (first entry)

DE Amino acid sequence of a raffinose synthase protein.

XX New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
XX
XX
PS Claim 27; Page 36-38; 55pp; English.
CC This sequence represents rapessed raffinose synthase, a protein
CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC hydroxy group attached to the carbon atom at the 6-position of the
CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
CC (see AA20210) encoding the enzyme was isolated from rapessed cv.
CC Westar leaf cDNA by PCR. Probes or primers generated from plant
CC raffinose synthase genes (see AA20207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
CC
XX
SQ Sequence 572 AA;
Query Match 90.1%; Score 508; DB 20; Length 572;
Best Local Similarity 87.1%; Pred. No. 3.4e-56;
Matches 88; Conservative 9; Mismatches 4; Indels 0; Gaps 0;
QY 1 PDMDMFQSTHPCAFHFAASRAISGGPIYVSDSVGNHFDLKLVPDGSILRSEYYALP 60
DB 304 PDMDMFQSTHPCAFHFAASRAISGGPIYISDCVGHDFDLRLRVLPDGSILRSEYYALP 363
QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVICAFNCGGWCRCR 101
DB 364 TRDRLFEEDPLHDGKTMKINLNKFTGIIIGAFNCGGWCRCR 404

RESULT 7
AAI32074
ID AAY32074 standard; Protein; 777 AA.
AC AAY32074;
XX
DT 17-JUN-2000 (first entry)
XX
DE Mustard raffinose synthase.
XX
KW Raffinose synthase; mustard; transgenic plant.
XX
OS Brassica juncea.
XX
FH Key Location/Qualifiers
FT Misc-difference 210 /note= "encoded by ACR"
XX
PN EP953643-A2.
XX
PD 03-NOV-1999.
XX
PE 27-APR-1999; 99EP-0107430.
XX
PR 30-APR-1998; 98JP-0120550.
XX
PR 30-APR-1998; 98JP-0120551.
XX
PR 04-DEC-1998; 98JP-0345590.
XX
PR 10-DEC-1998; 98JP-0351246.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Watanabe E, Oeda K;
XX
XX WPI; 1999-593144/51.
XX
DR N-PSDB; AA20209.
XX
PT New sense and antisense genes, useful for altering the level of

PT raffinose in food plants -
XX
XX
PS Claim 26; Page 29-31; 55pp; English.
CC This sequence represents mustard raffinose synthase, a protein
CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC hydroxy group attached to the carbon atom at the 6-position of the
CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
CC (see AA20209) encoding the enzyme was isolated from mustard
CC (Brassica juncea) leaf cDNA by PCR. Probes or primers generated from
CC plant raffinose synthase genes (see AA20207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
CC
XX
SQ Sequence 777 AA;
Query Match 89.7%; Score 506; DB 20; Length 777;
Best Local Similarity 87.1%; Pred. No. 9.4e-56;
Matches 88; Conservative 9; Mismatches 4; Indels 0; Gaps 0;
QY 1 PDMDMFQSTHPCAFHFAASRAISGGPIYVSDSVGNHFDLKLVPDGSILRSEYYALP 60
DB 509 PDMDMFQSTHPCAFHFAASRAISGGPIYISDCVGHDFDLRLRVLPDGSILRSEYYALP 568
QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVICAFNCGGWCRCR 101
DB 569 TRDRLFEEDPLHDGKTMKINLNKFTGIIIGAFNCGGWCRCR 609

RESULT 8
AAW57886
ID AAW57886 standard; Protein; 799 AA.
AC AAW57886;
XX
DT 23-SEP-1998 (first entry)
XX
DE Broad bean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
XX
KW gastrointestinal flora; broad bean.
XX
OS Vicia faba.
XX
FH EP849359-A2.
XX
PD 24-JUN-1998.
XX
PF 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Watanabe E;
XX
XX WPI; 1998-324670/29.
XX
DR N-PSDB; AAV40800.
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
XX
PS Claim 1; Page 26-29; 44pp; English.
XX
CC This sequence represents the broad bean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a

CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.

XX Sequence 799 AA:

Query Match 89.5%; Score 505; DB 19; Length 799;

Best Local Similarity 88.0%; Pred. No. 1.3e-55;

Matches 88; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 PDWDFQSTHPCAEFHAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 60

DB 529 PDWDFQSTHPCAEFHAASRAISGPIYVSDVGKHNFKLKSILVLPDGSILRQHYALP 588

OY 61 TRDCLFEDPLHNGETMLKIWNINKFTGVIGAFNCGGGMC 100

DB 589 TRDCLFEDPLHNGETMLKIWNINKYTGVLGLENCOGGGMC 628

RESULT 9

AAV30142

XX AAV30142 standard; Protein; 799 AA.

XX AAV30142;

XX 26-OCT-1999 (first entry)

XX Amino acid sequence of a raffinose synthase protein.

XX Raffinose synthase; plant; broad bean; sucrose; raffinose.

XX Victoria faba.

XX JP11215984-A.

XX 10-AUG-1999.

XX 12-DEC-1997; 97JP-0342899.

XX 28-NOV-1997; 97JP-0329006.

XX 18-DEC-1996; 96JP-0338673.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX WPI; 1999-511112/43.

XX N-PSDB; AA210001.

XX New raffinose synthase gene - is prepared from a plant material

XX Claim 5; Page 19-21; 40pp; Japanese.

XX The present sequence represents a raffinose synthase protein. The

XX sequence is isolated from plant material of broad beans. The

XX protein forms raffinose by complexing alpha(1 to 6) - D-galactosyl

XX hydroxyl group of the 6C of D-glucose residue in sucrose molecules.

XX Sequence 799 AA:

Query Match 89.5%; Score 505; DB 20; Length 799;

Best Local Similarity 88.0%; Pred. No. 1.3e-55;

Matches 88; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 PDWDFQSTHPCAEFHAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 60

DB 529 PDWDFQSTHPCAEFHAASRAISGPIYVSDVGKHNFKLKSILVLPDGSILRQHYALP 588

OY 61 TRDCLFEDPLHNGETMLKIWNINKFTGVIGAFNCGGGMC 100

DB 589 TRDCLFEDPLHNGETMLKIWNINKYTGVLGLENCOGGGMC 628

RESULT 10
AAB98659
ID AAB98659 standard; protein; 780 AA.

XX AAB98659;

XX 17-AUG-2001 (first entry)

XX Soybean protein; SEQ ID 1.

XX Mutant; mutein; raffinose synthase; raffinose oligosaccharide reduction;

XX plant; soybean.

XX Glycine max.

XX JP2001078783-A.

XX 27-MAR-2001.

XX 03-JUL-2000; 2000JP-0200571.

XX 09-JUL-1999; 99JP-0196036.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX WPI; 2001-313373/33.

XX N-PSDB; AAH27438.

XX Novel mutant protein of raffinose synthase is useful for reducing the

XX raffinose oligosaccharide content in a plant body -

XX Disclosure; Page 18-20; 30pp; Japanese.

XX The present invention relates to a mutant protein of raffinose synthase

XX in which at least one aromatic amino acid present at the position of

XX about 1-7 amino acids from the N-terminus is deleted or replaced. The

XX mutant protein can be used for reducing the raffinose oligosaccharide

XX content in a plant body. The present protein from soybean, was used in

XX the present invention.

XX Sequence 780 AA:

Query Match 88.3%; Score 498; DB 22; Length 780;

Best Local Similarity 86.0%; Pred. No. 1e-54;

Matches 86; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

OY 1 PDWDFQSTHPCAEFHAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 60

DB 510 PDWDFQSTHPCAEFHAASRAISGPIYVSDVGKHNFKLKSILVLPDGSILRQHYALP 569

OY 61 TRDCLFEDPLHNGETMLKIWNINKFTGVIGAFNCGGGMC 100

DB 570 TRDCLFEDPLHNGETMLKIWNINKYTGVLGLENCOGGGMC 609

RESULT 11

AAW57887

XX AAW57887 standard; Protein; 781 AA.

XX AAW57887;

XX 23-SEP-1998 (first entry)

XX Soybean raffinose synthetase.

XX Raffinose synthetase; metabolism modification; food additive;

XX gastrointestinal flora; soybean.

XX Glycine max.

XX EP849359-A2.

XX

PD 24-JUN-1998.
XX
XX 18-DEC-1997; 97EP-0122417.
PF
XX 18-DEC-1996; 96JP-0338673.
PR
XX (SUMO) SUMITOMO CHEM CO LTD.
PA
XX
XX
PI Oeda K, Wantanabe E;
DR
DR MPI: 1998-324670/29.
N-PSDB; AAY40801.
XX
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
XX
PS Claim 1; Page 31-34; 44pp: English.
XX
XX This sequence represents the soybean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
SQ Sequence 781 AA;

Query Match 88.1%; Score 497; DB 19; Length 781;
Best Local Similarity 85.0%; Pred. No. 1.4e-54;
Matches 85; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDDMDFOSTHPCAFHFAASRAISGPIYVSDSYGKHNFDLKTLVLPDGSILRSEYALP 60
DB 511 PDDMDFOSTHPCAFHFAASRAISGPIYVSDSYGKHNFKLKSLALPDGTLRCQHVALP 570
DB 571 TRDCLFEDPLHNGETMLKIWNLNKFTGVI GAFCNCGGWC 610
QY 61 TRDCLFEDPLHNGETMLKIWNLNKFTGVI GAFCNCGGWC 100
DB 571 TRDCLFEDPLHNGETMLKIWNLNKFTGVI GAFCNCGGWC 610

RESULT 12
AAY30143
ID AAY30143 standard; Protein; 781 AA.
XX
AC AAY30143;
XX
DT 26-OCT-1999 (first entry)
XX
XX Amino acid sequence of a raffinose synthase protein.
DE
XX
XX Raffinose synthase; plant; sucrose; raffinose.
KW
XX
XX Glycine max.
OS
XX
XX JP11215984-A.
PN
XX
XX 10-AUG-1999.
PD
XX
XX 12-DEC-1997; 97JP-0342899.
PF
XX
XX 28-NOV-1997; 97JP-0329006.
PR
XX 18-DEC-1996; 96JP-0338673.
PA
XX (SUMO) SUMITOMO CHEM CO LTD.
XX
XX MPI: 1999-511112/43.
DR
XX N-PSDB; AAY10002.
XX
PT New raffinose synthase gene - is prepared from a plant material

XX
PS Claim 8; Page 25-27; 40pp; Japanese.
XX
XX
XX The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
XX
SQ Sequence 781 AA;

Query Match 88.1%; Score 497; DB 20; Length 781;
Best Local Similarity 85.0%; Pred. No. 1.4e-54;
Matches 85; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDDMDFOSTHPCAFHFAASRAISGPIYVSDSYGKHNFDLKTLVLPDGSILRSEYALP 60
DB 511 PDDMDFOSTHPCAFHFAASRAISGPIYVSDSYGKHNFKLKSLALPDGTLRCQHVALP 570
DB 571 TRDCLFEDPLHNGETMLKIWNLNKFTGVI GAFCNCGGWC 610
QY 61 TRDCLFEDPLHNGETMLKIWNLNKFTGVI GAFCNCGGWC 100
DB 571 TRDCLFEDPLHNGETMLKIWNLNKFTGVI GAFCNCGGWC 610

RESULT 13
AAB49400
ID AAB49400 standard; Protein; 781 AA.
XX
AC AAB49400;
XX
DT 07-MAR-2001 (first entry)
XX
XX Soybean raffinose synthase.
DE
XX
XX Plant promoter; transgenic plant; desired property.
KW
XX
XX Glycine max.
OS
XX
XX EP1048733-A2.
PN
XX
XX 02-NOV-2000.
PD
XX
XX 27-APR-2000; 2000EP-0108962.
PF
XX
XX 30-APR-1999; 99JP-0124527.
PR
XX 01-SEP-1999; 99JP-0247211.
XX
XX (SUMO) SUMITOMO CHEM CO LTD.
PA
XX
XX Ishige F, Watanabe E, Oeda K;
PI
XX
XX MPI: 2001-104537/12.
DR
XX
XX N-PSDB; AAC89523.
DR
XX
XX New soybean plant promoters useful for generating transgenic plants
PT with desired properties -
PT
XX
XX Example 6; Page 24-27; 36pp; English.
PS
XX
XX The present invention provides novel plant promoters which can be used in
CC the production of transgenic plants which express genes with desired
CC properties.
CC
XX
XX
SQ Sequence 781 AA;

Query Match 88.1%; Score 497; DB 22; Length 781;
Best Local Similarity 85.0%; Pred. No. 1.4e-54;
Matches 85; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDDMDFOSTHPCAFHFAASRAISGPIYVSDSYGKHNFDLKTLVLPDGSILRSEYALP 60
DB 511 PDDMDFOSTHPCAFHFAASRAISGPIYVSDSYGKHNFKLKSLALPDGTLRCQHVALP 570
DB 571 TRDCLFEDPLHNGETMLKIWNLNKFTGVI GAFCNCGGWC 610
QY 61 TRDCLFEDPLHNGETMLKIWNLNKFTGVI GAFCNCGGWC 100

Db 571 TRDCLFEDPLHDGKTMKIKIWNLNKTYGVGLFNCGGGWC 610

RESULT 14

AA70978
ID AA770978 standard; Protein: 758 AA.

AA70978;

09-AUG-2000 (first entry)

Soybean raffinose synthase from clone sfl1.pk125.d4.

Soybean; raffinose synthase; raffinose saccharide;

clone sfl1.pk125.d4; nutritional; soy protein.

Glycine max.

MO200024915-A2.

04-MAY-2000.

22-OCT-1999; 99MO-US24923.

23-OCT-1998; 98US-0105451.

(DUPO) DU PONT DE NEMOURS & CO E. I.

Allen SM, Hiltz WD;

WPI; 2000-350754/30.

N-PSDB; AAD00335.

Nucleic acids and encoded proteins involved in the biosynthesis of

raffinose, useful for producing soybean seeds with a reduced raffinose

content and therefore improved nutritional quality -

Claim 2; Page 47-49; 58pp; English.

The present sequence is a raffinose synthase from

clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA

library sfl1. Raffinose synthase is involved in the biosynthesis

of raffinose and higher homologues in the raffinose saccharide family

from sucrose. The present sequence is useful for reducing the raffinose

saccharide content of soybean seeds which improves the nutritional

quality of the soy protein products derived from them.

Sequence 758 AA:

Query Match 87.6%; Score 494; DB 21; Length 758;

Best Local Similarity 84.2%; Pred. No. 3.2e-54;

Matches 85; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

1 PDMDFOGTHPCAEHAAASRAISGPIYSDVSGKHNFDLKKLVLPDGSILRSEYALP 60

487 PDMDFOGTHPCAEHAAASRAISGPIYSDVSGKHNFDLKKLVLPDGSILRSEYALP 546

61 TRDCLFEDPLHDGKTMKIKIWNLNKTYGVGLFNCGGGWC 101

TRDCLFADPLHDGKTMKIKIWNLNKTYGVGLFNCGGGWC 587

RESULT 15

AA732073

ID AA732073 standard; Protein: 783 AA.

AA732073;

17-JAN-2000 (first entry)

Sugarbeet raffinose synthase.

KW Raffinose synthase; sugarbeet; transgenic plant.

OS Beta vulgaris.

Key Location/Qualifiers

Misc-difference 227 /note= "encoded by CCW"

Misc-difference 348 /note= "encoded by CCR"

EP953643-A2.

03-NOV-1999.

27-APR-1999; 99EP-0107430.

30-APR-1998; 98JP-0120550.

30-APR-1998; 98JP-0120551.

04-DEC-1998; 98JP-0345590.

10-DEC-1998; 98JP-0351246.

(SUMO) SUMITOMO CHEM CO LTD.

Watanabe E, Oeda K;

WPI; 1999-593144/51.

N-PSDB; AA220208.

New sense and antisense genes, useful for altering the level of

raffinose in food plants -

Claim 25; Page 22-24; 55pp; English.

This sequence represents sugarbeet raffinose synthase, a protein

that can bind a D-galactosyl group through an alpha(1-6) bond to the

hydroxy group attached to the carbon atom at the 6-position of the

D-glucose residue in a sucrose molecule to form raffinose. cDNA

(see AA220208) encoding the enzyme was isolated from sugarbeet cv.

harming leaf cDNA by PCR. Probes or primers generated from plant

raffinose synthase genes (see AA220207-10) may be used to obtain

other raffinose synthase genes by labeled detection or amplification

(claimed). These genes may be used to control the levels of

raffinose produced in plants. Antisense genes can be used to knock

out existing gene activity, and sense genes to increase the level

of gene activity. The resulting transgenic plants may be used as a

food source to alter the growing conditions for gut enterobacteria,

providing general health advantages.

Sequence 783 AA:

Query Match 84.4%; Score 476; DB 20; Length 783;

Best Local Similarity 83.2%; Pred. No. 6.8e-52;

Matches 84; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

1 PDMDFOGTHPCAEHAAASRAISGPIYSDVSGKHNFDLKKLVLPDGSILRSEYALP 60

510 PDMDFOGTHPCAEHAAASRAISGPIYSDVSGKHNFDLKKLVLPDGSILRSEYALP 569

61 TRDCLFEDPLHDGKTMKIKIWNLNKTYGVGLFNCGGGWC 101

TRDCLFADPLHDGKTMKIKIWNLNKTYGVGLFNCGGGWC 610

Search completed: March 27, 2003, 10:03:45

Job time : 72.439 secs


```
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 266
; LENGTH: 1138
; TYPE: PRT
; ORGANISM: Shigella flexneri
US-10-043-487-266

Query Match
Best Local Similarity 26.5%; Pred. No. 27;
Matches 26: Conservative 13; Mismatches 27; Indels 32; Gaps 5;

QY 1 PWDMEFQSTHPCAFHNASRAISGPIYSDSV-----GKNHFDLKL----- 44
DB 691 PFDLNLQNLH-----GSLVQDVVEYLESTVRPRVSLQDDYFATLRELENTLRTQSL 743

QY 45 ---VLPGDSLIRSEYALPTRDCLFEDPLHNGETMLKI 79
DB 744 SLAVIPEGKIMNNYY---QECLEFY--LHNYSTNLAI 775

RESULT 6
US-09-801-368-148
; Sequence 148, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801.368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 148
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-148

Query Match
Best Local Similarity 11.4%; Score 64.5; DB 10; Length 358;
Matches 24: Conservative 10; Mismatches 24; Indels 33; Gaps 4;

QY 15 FHAAASRAISGPIYSDSVGKNHFDLKLVLPGDSI-----LRSEYAL-----PTRD 63
DB 186 FNA-----YGTANFLQENFEAIKKVYVCPDSCLOHQRPKRPRKRSLLLLIPNMS 235

QY 64 CLFEDPLHNGETMLKIWNLKNKFTGVGATNC 94
DB 236 ELSETP-----LMRFAGVGEFMNC 254

RESULT 7
US-10-047-539-2
; Sequence 2, Application US/10047539
; Patent No. US20020177547A1
```

```
; GENERAL INFORMATION:
; APPLICANT: MOLLING, KARIN
; APPLICANT: PAVLOVIC, JOVAN
; APPLICANT: NAMRATH, MICHAEL
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS FOR TREATING OR PREVENTING
; FILE REFERENCE: VOS-27
; CURRENT APPLICATION NUMBER: US/10/047,539
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: EP 01 10 0914.9
; PRIOR FILING DATE: 2001-01-16
; SOFTWARE: PatentIn Ver. 2.1
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO: 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-047-539-2

Query Match
Best Local Similarity 11.4%; Score 64.5; DB 9; Length 626;
Matches 29: Conservative 17; Mismatches 42; Indels 29; Gaps 7;

QY 1 PWDMEFQSTHPCAFHNASRAISGPIYV-----SDSVGKNHFDLKLVLPGDSIL----- 52
DB 50 PEMTEVQSGNCWRGQVSLRVINDGPTLVGANASFIALH-FPGSOK-VLPDGOVIMANN 107

QY 53 -----RSEYALPTRDCLFED--PLHNG-----ETMLKIWNL-NKRTGVIG 90
DB 108 TIINSQVWGQPYRQEPDDACVFPDGGPCPSGPRKPRRSFVYWKTKGKYQVIG 164

RESULT 8
US-10-081-281-93
; Sequence 93, Application US/10081281
; Patent No. US20020151707A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Gross, Jane A.
; APPLICANT: Sheppard, Paul
; TITLE OF INVENTION: Immune Mediators and Related Methods
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/081,281
; FILING DATE: 20-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/261,811A
; FILING DATE: 03-Mar-1999
; APPLICATION NUMBER: US 08/480,002
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/482,133
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/483,241
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 60/005,964
; FILING DATE: 27-OCT-1995
; APPLICATION NUMBER: US 08/657,581
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:07:51 ; Search time 41 Seconds
(without alignments)
3940.023 Million cell updates/sec

Title: US-09-675-208-5

Perfect score: 4225

Sequence: 1 MAPSFKNGSNVVSFDGLND.....QVWPIDSSGCSGIVIEYLF 784

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mnc:*
- 8: sp-organella:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-virus:*
- 16: sp-bacteriap:*
- 17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4325	100.0	784	10 Q92T62	Q92T62 cucumis sat
2	2945	69.7	783	10 Q9FND9	Q9FND9 arabidopsis
3	2807.5	66.4	798	10 Q8VWV6	Q8VWV6 pisum sativ
4	2643	62.6	816	10 Q942N6	Q942N6 oryza sativ
5	1914.5	45.3	853	10 Q93XK2	Q93XK2 pisum sativ
6	1914	45.3	857	10 Q9S8Z0	Q9S8Z0 phaseolus a
7	1868.5	44.2	863	10 Q8RW08	Q8RW08 strachys aff
8	1853.5	43.9	807	10 Q9SXY4	Q9SXY4 arabidopsis
9	1482	35.1	757	10 Q40077	Q40077 hordeum vul
10	1473.5	34.9	765	10 Q43408	Q43408 brassica o1
11	1459.5	34.5	773	10 Q94A08	Q94A08 arabidopsis
12	1458.5	34.5	773	10 Q9SCM1	Q9SCM1 arabidopsis
13	1446.5	34.2	844	10 Q8RX87	Q8RX87 arabidopsis
14	1413.5	33.5	779	10 Q9M4M7	Q9M4M7 persea amer
15	1140	27.0	1170	10 Q9LFF7	Q9LFF7 arabidopsis
16	697.5	16.5	386	10 Q9M442	Q9M442 cicer ariet

17	514.5	12.2	283	10 Q9FWM2	Q9FWM2 arabidopsis
18	490.5	11.6	357	10 Q94666	Q94666 cicer ariet
19	471.5	11.2	371	10 Q04607	Q04607 arabidopsis
20	408.5	9.7	702	2 Q93CM6	Q93CM6 bifidobacte
21	397	9.4	674	17 Q96X62	Q96X62 sulfolobus
22	365	8.6	649	17 Q97094	Q97094 sulfolobus
23	350	8.3	125	10 Q42099	Q42099 arabidopsis
24	219	5.2	104	10 Q8W2G5	Q8W2G5 oryza sativ
25	161	3.8	204	10 Q9AV63	Q9AV63 oryza sativ
26	145	3.4	88	10 Q42092	Q42092 arabidopsis
27	121.5	2.9	538	16 P74506	P74506 synechocyst
28	114.5	2.7	776	2 Q8RM03	Q8RM03 xanthobacte
29	114	2.7	623	5 Q61391	Q61391 caenorhabdi
30	114	2.7	623	5 Q95ZJ2	Q95ZJ2 caenorhabdi
31	113.5	2.7	632	3 Q04049	Q04049 saccharomyc
32	112.5	2.7	1963	10 Q94HY2	Q94HY2 oryza sativ
33	111	2.6	741	10 Q9LNO1	Q9LNO1 arabidopsis
34	106	2.5	4074	4 Q8TC29	Q8TC29 homo sapien
35	105.5	2.5	473	16 Q8XM38	Q8XM38 clostridium
36	105.5	2.5	568	3 Q9P8X6	Q9P8X6 aspergillus
37	104.5	2.5	736	16 Q92M14	Q92M14 rhizobium m
38	104	2.5	482	10 Q65328	Q65328 nicotiana g
39	104	2.5	689	17 Q976H7	Q976H7 sulfolobus
40	104	2.5	933	10 Q8WSR2	Q8WSR2 arabidopsis
41	103.5	2.4	1896	12 Q04232	Q04232 mucosal dis
42	103	2.4	3944	5 Q18667	Q18667 caenorhabdi
43	102.5	2.4	544	16 Q98RA7	Q98RA7 mycoplasma
44	102.5	2.4	582	16 P73732	P73732 synechocyst
45	102.5	2.4	1093	10 Q9LEP6	Q9LEP6 arabidopsis

ALIGNMENTS

RESULT 1

ID Q92T62 PRELIMINARY: PRT: 784 AA.

AC Q92T62;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Raffinose synthase (EC 2.4.1.82).

DE RFS.

GN Cucumis sativus (Cucurbit).

OS Cucumis sativus (Cucurbit).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC Cucurbitales; Cucurbitaceae; Cucumis.

OX NCBI_TaxID=3659;

OX [1]

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LEAF;

RA Ohsumi C., Nozaki J., Kida T.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF073744; AAD02832.1; -;

KW Glycosyltransferase; Transferase.

SQ SEQUENCE 784 AA; 86920 MW; 3806A491F0908933 CRC64;

QY	1	MAPSFKNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVPENIVASPPYTSIDKS	60
DB	1	MAPSFKNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVPENIVASPPYTSIDKS	60
QY	61	PVSVCEVGFDASEPDSSRHVSIKGLKDIRFMSIFRFKVMWTHWVGRNGDLESTQIV	120
DB	61	PVSVCEVGFDASEPDSSRHVSIKGLKDIRFMSIFRFKVMWTHWVGRNGDLESTQIV	120
QY	121	ILEKSDSGRPVFLPIVGGPRTSTQPCDDDFVQVVCSSGSKVVDASFRSMVLTHAGD	180
DB	121	ILEKSDSGRPVFLPIVGGPRTSTQPCDDDFVQVVCSSGSKVVDASFRSMVLTHAGD	180

OY	181	DPFALVKEAMKIVTRHLCSTFTLLSEKKPPGVLDVFSGCTMDAFLTVNPGVIEGVAHLV	240
Db	181	DPFALVKRAMKIIVTRHLCSTFTLLSEKKPRGLVDKFGMCTMDAFILTVNPGVIEGVAHLV	240
OY	241	DGGCEPGLVLIDDGWQSIIGHSDSPITREGMNQTVAGEOMPRLIKFEQENTKFRDYVNPKA	3000
Db	241	DGGCEPGLVLIDDGWQSIGHSDEPTREGMNQTVAGEOMPCRLIKFOENKFKRDYVNPKA	3000
OY	301	TGPRAGOKGMAKFIDEELKGEEKTYEHYVNHALLCGWGMYGARPOVPSGLEARVIOPVLSPG	3660
Db	301	TGPRAGOKGMAKFIDEELGEEKTYEHYVNHALLCGTWGGLLEPROVPSGLEARVIOPVLSPG	3660
OY	361	LQMTEMEDLAVDKIYLHKVGLVPPEKAEMEYEGSLAHLEKVGIIDGVKIDVLIHLLMLCEDY	4200
Db	361	LQMTEMEDLAVDKIYLHKVGLVPEKAEMEGSLAHLEKVGIIDGVKIDVLIHLLMLCEDY	4200
OY	421	GGRVDLAAYYKAKMTKTSINKHFKNQGYASAEHCNDWMFLGTALISIGRGDDFWCWDPS	4800
Db	421	GGRVDLAAYYKAKMTKSINKHFKNQGYASAEHCNDWMFLGTALISIGRVGDDEFWCWDPS	4800
OY	481	GDPNGTEFWLGCCHHVHCANDSLMANGNIHPDMDFOSTHPCAFAHAASRAISGGPIYVD	5400
Db	481	GDPNGTEFWLGCCHHVHCANDSLMANGNFIHPDMDFOSTHPCAFAHAASRAISGGPIYVD	5400
OY	541	SVGHNPDLKLKLVLPDSILRSEYIALPTRDCLEFDPLHNGETMLKTIMLNKFTGYGA	6000
Db	541	SVGHNPDLKLKLVLPDSLIRSEYIALPTRDCLFEDPLHNGETMLKTIMLNKFTGYIGA	6000
OY	601	FNCGGGCMCFETRNNOCSSOYSKYVTAKTNPKDIEMHSGENPISIECVTFEVALYLYOAK	6600
Db	601	FNCGGGCMCFETRNNQCSSOYSKYVTSKTNPKDIEMHSGENPISIECVKFPALXLYOAK	6600
OY	661	LILSKPSODLDIALDPEEFELLIVSPVTKLIQTSLHFAIPGLVNMLTSAIGALSVDYDD	7200
Db	661	LILSKPSODLEDIALDPEEFELLIVSPVTKLIQTSLHFAIPGLVNMLTSAIGALSVDYDD	7200
OY	721	LSVSIEIGKGGCKEMRPVASKRKRCRIDGEDVGKXYDDOMVVQVWPIDSSSGGISVI	7800
Db	721	LSVSIEIGKGGCKEMRPVASKRKRCRIDGEDVGKXYDDOVVVQVWPIDSSSGGISVI	7800
OY	781	EYLF 784 IIII	
Db	781	EYLF 784	
<hr/>			
RESULT 2			
O9FEND9		PRELIMINARY;	PRT; 783 AA.
AC O9END9			
DT 01-MAR-2001 (TREMBlrel. 16, Created)			
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)			
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)			
DE Ratifinose synthase protein.			
GN MP012.13.			
OS Arabidopsis thaliana (Mouse-ear cross).			
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;			
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots; Rosidae:			
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.			
OX NCBI_TaxId=3702;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=COLUMBIA;			
RX MEDLINE=98069011; PubMed=9405937;			
RA Kocani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,			
RA Tabata S.;			
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.			
RT Sequence features of the regions of 1,044,062 bp covered by thirteen			
RT physically assigned P1 clones."			
RNA Res. 4:291-300(1997).			
RN [2]			
RP SEQUENCE FROM N.A.			
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,			
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,			

QY	DB	Query Match	Best Local Similarity	69.7%;	Score 2945;	DB 10;	Length 783;
		Matches 544;	Conservative 99;	Mismatches 117;	Indels 26;	Gaps 9;	
QY	14	SFDDGLN--DMSSPRAIDSDFTVNGHSLSDVPEVIVASPSPTYSIDKS----	PYSVCGF	67			
DB	9	SDSSINGVDFFKEFRLEEDSTLLANQVVLTPVAVVLTSSPVL--VDDGVPPLDVSAGSF	67				
QY	68	VGFLA--SEPDSRHVVSTKLDKIDRFMSIFERKVVMTTHWVRNGDGLSENOITLEK--	124				
DB	68	IGFNLDSGRKHHVASICKLNIRMSIFERKVVMTTHWVSNGDIDENQIILLDSSG	127				
QY	125	-----SDSGRPVFLPIVEGPRFTSTIQPDDEFDVDCVSSGSKVVDASFRLMLYLA	178				
DB	128	SDSGRPSGSGSRPVYLLPRLLEGSRFSRQSGEDDVAACVSGSRETVTSGEFRQIVYVHA	187				
QY	179	GDDEPALKEAMKIVTRPLTGFRLLEEKTPPGIVDKFEMCTWDATYLVVHOGVLEGRA	238				
DB	188	GDDEPKLKYDKAMKIVRYMNTFKLLEEKSPGIDYKFEWCTWDATYLVVNEGVIKGYKC	247				
QY	239	LVDSGCPRLVLLIDGWOSTGHDSDPIITKEGMMQTVAGEOMFCRLKFEQENKFRDYVP	298				
DB	248	LVDSGCPRLVLLIDGWOSTGHDSDPIITKEGMMQTVAGEOMFCRLKFEQENKFRDYVP	307				
QY	299	KATGRPAGQKMKAFIDELKGEFTVEVYVWHALCGTWGSLRPVQRLPEARYIQPLS	358				
DB	308	K----DONDVGMKAFVRRLDKDEFSTVDIIVYHALCGTWGSLRPAPRLPSTIIRPELS	363				
QY	359	PGLQMTMEDLAVDKITVLHKVLPPEKAEEMVEGILHAILEKVIDGKVLDIYHILEMICE	418				
DB	364	PGLKLTMDLAVDKITIEIGTIFASPDIAKFEYEGHSHLQNAAGIDGVVDYHILEMICO	423				
QY	419	DYSGRVDLAKAYYKAMTKSINKHFKGNGVIASMEHNDPFLGLTEALSILGRVDDFWCTD	478				
DB	424	KYSGRVDLAKAYYFALTKTSYVKHFNNGVIVIASMEHNDPFLGLTEALISILGRVDDFWCTD	483				
QY	479	PSGPNPGEFMLOGGHVHCANDSLMNGFTHPDMQPOSTHPCAFHNASRAISGCPYV	538				
DB	484	PSGPNPGEFMLOGGHVHCAYNSLMNGFTOPDMQPOSTHPCAFHNASRAISGCPYI	543				
QY	539	SDSYGKHNFDELKLKVLIPDGSILRSEYVALPTDCLFEDPRLANGFTMLKINLNKFTGVI	598				
DB	544	SDCYGKHDFELKLKVLIPDGSILRSEYVALPTDCLFEDPRLANGFTMLKINLNKFTGVI	603				
QY	599	GAFNCGGGGWRERTRRNOCFSECVNTLATTSPKDVENSSSPSTIANVVEEFALELSOS	663				
DB	604	GAFNCGGGGWRERTRRNOCFSECVNTLATTSPKDVENSSSPSTIANVVEEFALELSOS	663				
QY	659	KLLILSKPSODLDALDPFEFELITVSPVTKLITSLFFAPIGLVNMLNTSGAIDQSVYD	718				
DB	664	KLLILSGINDLELTLEPFKELITVSPVYITIEGNSVFAFATIGLVNMLNTSGAISRSLYVN	723				
QY	719	DDLSSVEIGKCGCEMRVAFASKPRACRIDGEDVGFYKIDODOMVAVVWPBIDSSSGIS	778				

```

Db 724 DE--SVENGVFAGGEFRYVASKKPVSCILIDGVEVFGY-EDSMVWVQVPM---SGPDGLS 777
Oy 779 VIEYLF 784
Db 778 STQYLF 783

RESULT 3
OBYWN6 PRELIMINARY: PRT: 798 AA.
AC OBYWN6.
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Raffinose synthase (EC 2.4.1.82).
GN RFS.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxId=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WUNDER VON KELDON; TISSUE=SEED;
RA Peterbauer T., Mach L., Mucha J., Richter A.;
RT "Molecular characterization of raffinose synthase from pea (Pisum
RT sativum L.) seeds."
RU Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ264475; CAD20127.2; -.
KM Transferase; Glycosyltransferase.
SQ SEQUENCE 798 AA; 88717 MW; 8D3F3ED5B8617B7 CRC64;

Query Match 66.4%; Score 2807.5; DB 10; Length 798;
Best Local Similarity 65.1%; Pred. No. 7.8e-220;
Matches 510; Conservative 120; Mismatches 125; Indels 29; Gaps 10;

Oy 22 SSP---FAIDGS-DFTVNGHSLSDVPENI---VASPSPTSIDKSPVSV----- 64
Db 23 NSPLSLISIDGSRNFLVGNHPFLTGVPRIITTTTSTSPPLDFKSNKNTIANNNNTLQ 82
Oy 65 -GCFVGFQASPDSDRHVVSIGKLDIRFMSIFRFKVMWTTTHVGRNGGLESFQVILE 123
Db 83 QCFVGFQFNTTEAKSHHVPLGKLGKIKFTSIFRFKVMWTTTHVGRNGHLEHOTLILD 142
Oy 124 KSDS--GRPVFLPTIYEGFRTSIQDQDDFVQVCESSGSSKVVQASFRSMYLHAGDP 182
Db 143 KVISLGRPIVLLPLIENSFTSLQGLNDYDMSVSGSTHVTGSTRFACLYLHLNDP 202
Oy 183 FALVKEAMKIVRTHLGTFRLEEKTPPGIIVDKFGMCTMDAFYLVAPQGVIEGRHLVDG 242
Db 203 YRLVKEAVAVIOTKLGTFLEKTPPSLIEKFGMCTMDAFYLVAPQGVIEGRHLVDG 262
Oy 243 GCRPGLVILIDGQWOSTIGH-DSDPIT--KEGMOIVAGEOMPCRLKFOENYKFRDYNPKA 300
Db 263 GCRPGLVILIDGQWOSTISHDDDDPVERDGMNRTSAGEQMPCLIKKEENYKFRVEN-- 319
Oy 301 TCRPAGOKGMAKAFIDELGEEFKTEVHVYVHMLCGYWGGLRPOVPLPARVYIOPVLSG 360
Db 320 -GDNCKKSLVGVFRDLKKEEFSVESVYVHMLCGYWGGLRPOVPLPARVYIOPVLSG 378
Oy 361 LQMTMEDLAVDKIVLHKVGLVPRKEAEMYEGSLHAHLEKVIDGVKIDYIHLEMLECDY 420
Db 379 VKMTMEDLAVDKIVENGVLVPRNLQEMFDSHLSAGIDGVVDYIHLELLESEY 438
Oy 421 GGRVLDAAKYTAAMTSIKKHKGNVYASMEHNDFMFLGTEAISLGVAGDDFQCTDSS 480
Db 439 GGRVLDAAKYTAALSSVKKHFGKNGVYASMEHNDFFLLGTEAISLGVAGDDFQCTDSS 498
Oy 481 GPRNGTFMLOGGCHMYCANDSLMGNFIHPDMDMFOSTPRCAAFHAAASAIISGPIYVD 540
Db 499 GPRNGTFMLOGGCHMYCANSLSMGNFIHPDMDMFOSTPRCAEFHAAASAIISGPIYVD 558
Oy 541 SVGKHNFDDLKLVLPDGSILRSEYVALPTRDCLFEDPLHNGETMLKIVNLNKFVGIGA 600

```

```

Db 559 CVGNHNFKLKLSFVLPDGSILRQHYALPTRDCLFEDPLHNGKTMKIVNLNKFVAGVGL 618
Oy 601 FNCGGGCMCRETRRNOCSQVSKRTSKNPKDIMHSGENISTEGYTFALYLOAKK 660
Db 619 FNCGGGCMCRETRRNOCSQVSKRTSKNPKDIMHSGENISTEGYTFALYLOAKK 678
Oy 661 LILSKPSODLDIALDPFEEBELITVSPVTKLQTSLSHFAPIGLVNMLNTSGAIOQVDD 720
Db 679 LSLMKCSORLEVSLEPFSELMTVSPKLVFYSKRLQFAPIGLVNMLNTSGAIOQVDD 738
Oy 721 LSSVEIGVKGCGEMRVFASKPRACRIDGEDVGFKYDDQMVVYVPMPISSSGTISY 780
Db 739 ASLVKIGVAGCGELSVFASERPVCKIDGVSEFDP-EDKVVRYQILMP---GSSPLSLV 794
Oy 781 EYLF 784
Db 795 EYLF 798

RESULT 4
O942T6 PRELIMINARY: PRT: 816 AA.
AC O942T6.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Putative raffinose synthase.
GN P0583G08.2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxId=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0583G08."
RU Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003282; BAB64768.1; -.
KM Transferase.
SQ SEQUENCE 816 AA; 89588 MW; B316EDF3566C5178 CRC64;

Query Match 62.6%; Score 2643; DB 10; Length 816;
Best Local Similarity 59.6%; Pred. No. 2.1e-206;
Matches 500; Conservative 111; Mismatches 150; Indels 78; Gaps 16;

Oy 1 MAPSEKNG-----SNVSEFDGLNDSSP--FAIDGSDFTVNGHSLSDVPENIVASPSPT 55
Db 1 MAPNLSKAQDLDLIGVVAVDGL--IKPRFTLKGLDVAIDGHPFLLDVAPNIRLTPAS-T 57
Oy 56 SIDKSPV---SVGCFVGFQASPDSDRHVVSIGKLDIRFMSIFRFKVMWTTTHVGRNGD 112
Db 58 LVPNSDVPAALGAFSLGFDAPAKRHHVPIKLRDTRFMSIFRFKVMWTTTHVGRNGD 117
Oy 113 LSESDQVILEKSDS-----GRPVFLPTIYEGFRTSIQDQDDFVQVCESSGSSKVV 165
Db 118 VENEIOMMLDSDSGTKSSFTGPRPIVLLPIYEGFRTSIQDQDDFVQVCESSGSSKVV 177
Oy 166 VASFRSMYLVLAGDDPFAVLEAMKIVRTHLGTFRLEEKTPPGIIVDKFGMCTMDAFYL 225
Db 178 RGSVFRSAAYLVHAGDDPFLVDARVVRANHLGTRFLMEKTPPGIIVDKFGMCTMDAFYL 237
Oy 226 TVHPQGVIEGRHLVDGCGCRPGLVILIDGQWOSTIGHSDPIT--KEGMOIVAGEOMPCRL 283
Db 238 KYHPCEVWEGVRRLADGCGCRPGLVILIDGQWOSTICHDDDLGSCAEGMNTSAGEQMPCL 297
Oy 284 LKFOENYKFRDYNVNRKATSPRAGOKMAKAFIDELGEEFKTEVHVYVHMLCGYWGGLRPO 343
Db 298 IKFOENYKFRY-----KGGMGTFVRDKMAKAFPTVDOYVYVHMLCGYWGGLRPO 346
Oy 344 VPGLPARVIOVPLSPGLQMTMEDLAVDKIVLHKVGLVPRKEAEMYEGSLHAHLEKVID 403

```


Query Match 45.3%; Score 1914; DB 10; Length 857;
Best Local Similarity 43.8%; Pred. No. 5,8e-147;
Matches 370; Conservative 139; Mismatches 243; Indels 92; Gaps 10;

17 GUNDMSSPAIDGSDFTVNGHSLSDVPENIYASPSPTST---DKSPSYV----- 64
13 GLEPEKEVFDLSDGKLTGVKGVLLSHVPEVNTF---SFSISICVPRADPSSIIQRYTAASH 70
65 -CCFVGFQASDSEDSRHVYISGKLKDIRFMSIFRFKVMFTTHVGRNGGLESETQVILE 123
71 KCGFLGFSVNSDRLINLSFGRGNFISTIRFTWSTONVNGSGLQMETOWILIE 130
124 KSDSGRPVFLPIVEGPFRTSIOPGDDFVDCVESGSSKVVADSFMSLYLHAGDDPF 183
131 VPEI-ESYVVIPIIEKFSRSLAHGSDDHVKAICAESSTQVIRASSFGALVAVHAETPY 189
184 ALVKEAMKIVRNHLGTFRLLEKTPPGIYDKFGMCTWDAFYLYVHPQVIEGVRLVDCG 243
190 NLMREAYSLARVHLDSFRLLEKTPPRIYDKFGMCTWDAFYLYVNPVNGVWHLGKDFSEGG 249
244 CPGGLVLIDGMSQISGHOSDPTKEGMSNOTVAGEOMPCLLKFOEYKFRDYVNPKATGP 303
250 VAPRFVYIDGMSQVNFDEDEPNEDAKNLVLCGEQMTARLHFEESDKFRKYQKGLLGP 309
304 RA-----GQK-- 308
310 NAFSPNETIKELKIGTAEHLGKQAAIASGSDLAIELMYKVRREIDLEGCKG 369
309 -----GMAKFIDELKEKFTYVEHYVWHALCGYWGRLRPQVPGLEAPRY 352
370 ESNEGGCCCAAECCGMDFTTDLRTKEFGKLDVYVWHALCGMGVPRGTTHL-DSKX 428
353 IOPVLSPGLOMTMEDLAVDKIVLHKVGLVPRPEKAEEMVGLHAHLEKVIDGVKIDVIL 412
429 ITCRLSPGLVGTMKDLANDKIVEGSLGVHPQANDLDSMSHLYAQVIGVTKIDVILHS 488
413 LEMLCEDYGGVRLAKAAYKAMTKSINKHFKNGVYASMEHNDPMFLGTEAISLGRVD 472
489 LEYVEEYGGVRAELKAYYDGLTNSIKNFNGSGLIASMOGNDFFLGTCKQIPRGVSD 548
473 DWCHDPSGDPRGTFWLQCGHNVHCANDSLAMGNFTHPMDMFQSTHPCAAHAAHSAIS 532
549 DMFQDPPNDPGVFWLQCGHNVHICSYSLMGMGOIOPDMDMFQSDHECAKFNAGSRAIC 608
533 GGPVIVSDVSGKHNFDLKLKLVLPDGSILRSEYVALPTPDCLEFPEPLHNGEMTKMKNIN 592
609 GGPVIVSDVSGHDEDLKLKLVPRDGTVPKCYFLPLTRDCLFRNPLFDQKTVLTKIMFN 668
593 KFTGVIGAFNCGGMCRETRRNOGFSQYSKRVTSKTNPKDIEHMSGENDPISIEGVKTPA 652
669 KYGVYIGAFNCGGAMDPRGKKFKGPRPECKYKALISCVHVTREVMQOKKEAEHMGKAEVY 728
653 LYLTAOKKILISKP-SQDLIDALDPEFEFLTVSPYTKLIQTSLHAPRIGLYNMMLTSGA 711
729 VYLNDAEVLHMLTPVSEPLQTLTOPSTFELYNFVPEKSGSSINIKFAPIGLTNMFNSGCT 788
712 IOSVDYDDLSEVETGVGCGEMRPAKCRIDGSEVGFKKDDQDMVVVQVWPMPID 771
789 IOLELEYE--KDYKVKVKGGRFLAYSTOSPCKFQOLNNGDAAFOWLPDGLTKLTLNLAW-IE 845
772 SSSG 775
846 ENDG 849

RESULT 7
Q8RM08 PRELIMINARY: PRT: 863 AA.
AC Q8RM08:
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Stachyose synthase (EC 2.4.1.67).

GN STS.
OS Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;
OC Eukaryota: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
OC Asteridae: euasterids I; Lamiales; Lamiaceae; Stachys.
OX NCBI_TaxId=168825;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Pesch M., Schmitz K.;
RT "Molecular cloning of a cDNA encoding for stachyose synthase from
RT Stachys sieboldii."
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ344091; CAC66963.1; -.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 863 AA; 95227 MW; D96F666090DEPD CRC64;

Query Match 44.2%; Score 1868.5; DB 10; Length 863;
Best Local Similarity 42.3%; Pred. No. 3e-143;
Matches 366; Conservative 143; Mismatches 253; Indels 103; Gaps 9;

16 DGLNDKSP-----FALDGSFTVNGHSLSDVPENIV-----ASPSPT 55
6 DPLISSIFSLISVKKDNAAELVGGKLSYKNNVPLSEIPENVTFKSSICQSSGAPAPLY 65
56 SLDKSPSVYGCVPDASPDSDRHVYISGKLKDIRMSIFRFKVMFTTHVGRNGGDL 115
66 NRAQSLNSCGFLGFSQKESADSVNSLGRFTNRREVSTIRFTWSTONVNGSGLQ 125
116 ETQIVLIEKSDSGRPVFLPIVEGPFRTSIOPGDDFVDCVESGSSKVVADSFMSLY 175
126 ETQWIMLNPEI-KSYAVVPIVEGKFSRSLAHGSDDHVKAICAESSTQVIRASSFG 184
176 LHAQDDPFLVKEAMKIVRNHLGTFRLLEKTPPGIYDKFGMCTWDAFYLYVHPQVIEG 235
185 VHVSDMPYTLKMDGYTAHVHLDTEFLIEKSPAPLVNFKFGMCTWDAFYLYVHPQVIEG 244
236 VRLVHVGCGPGLVILIDGMSQISGHOSDPTKEGMSNOTVAGEOMPCLLKFOEYKFR 295
245 VKEFSDGGSFPAFLIIDGMSQINIDGQDPNEDAKNLVLCGEQMTARLHFEESDKFR 304
296 VNPKATGP----- 303
305 KGGSMGPRVYFDPKPRLLISKATEIEGVKARDKALIOSGITDLSQYEIKKLKINKEL 364
304 -----RAGQKMAKFIDELKEKFTYVEHYVWHALCGYWGRL 340
365 DEMFGGNDGKSSKGCSDCSQNSGSKAFTNDLRNFKGLDIDYVWHALAGAMGV 424
341 RPOVPLPEARVIOPLSPGLQMTMEDLAVDKIVLHKVGLVPRPEKAEEMVGLHAHLEK 400
425 KPGATHL-NAKLEPCKLSGLDGTMTDLAVVKILBESISLVHFDQAEEDYDSHISTLSV 483
401 GIDGVKIDVILHLEMLCEDYGGVRLAKAAYKAMTKSINKHFKNGVYASMEHNDPMFL 460
484 GITGVKVDVYIHLLEYVSEYVGGVRELKAYYKGLSLLKNNNGSGLISSMOGNDFFLL 543
461 GTEALISLGVAGDFFWCTDPSGDPNGTFLWQCGHNVHCANDSLAMGNFTHPMDMFQST 520
544 GTEOISMGRVGDFFWQDNDGPMGVFWLQGVHIMICAAVNSMMGQIIPDMDMFQSDHC 603
521 CAAPHAASRAISGSPVYSDVSGKHNFDLKLKLVLPDGSILRSEYVALPTPDCLEF 580
604 SAKFNAGSRAICGGPVYSDSLGHDFFDLKLKLVFPGCTIIPKCIHPLPTRDLFLFNPLF 663
581 NGETMLKIMNLKFTGVIGAFNCGGMCRETRRNOGFSQYSKRVTSKTNPKDIEHMSG 640
664 DSKTIKIMNFKYGVVAFNCGGAMPKQRIKGYSECKKPLSGSVHVDIENDQV 723
641 NPISIEGVKTFALYLOAKKILSKPSQD-LDIALDPEFEFLTVSPYTKLIQTSLHAP 699
724 EATKMGAEVAVVYLSESKLLTLTPESDPIPTFTLKSTFEIFSPVPIKKLGO-GVKFAP 782

Db 125 LRLLEGGRVTLQGDNDDELQICIESGDKAVETEGGMNNVYVNAGTNPFDTITQAVKAVE 184

QY 195 TRLGTFRLLEKTRPGYIDKFGMCTWDAFYULVTRPOGYIEGVRLHLDGCGRRGLVLDG 254

Db 185 KHTQTFHREKKTTPSPFEDMGCTWDAFYDVTADGCKOGLRSIABEGAPRFLITIDG 244

QY 255 WOSIGHDS -BPIRKEGNQNOTVAGROMCRLKLPDENKCFRDYVPAKRTGRRAGCKMKA 312

Db 245 WQOIGSEKKDP - - -GV-AVOEGAGFASRLTGIREFNKFGSEHNQET - - - - -PGLKR 293

QY 313 FIDELGKFEKTVENHYVNALCGYWGAGRPVPGSLR - - -EARYIDPVLSPGLQMTMED;AV 370

Db 294 LVDETKKE -HGKSYVYVNAHAGYWGKVRPSAGCHENEPALATVYVSPGYTGNPDIYM 352

QY 371 DKVLHKKVGLVPRKEAEEMVEGLNHLNLEKVGIDGVKIDVILHLEMLCEDYGGRYDLAKAY 430

Db 353 DLSLVGLGVLNPRVHREFYDELNAYLACGVGVKVDQVQNIETLGAHGGRVALTRAY 412

QY 431 YKAMTKSLNKKFPGNGVYASMEHCNDFLGTETASISLRGVDDDFMCTPDSGDPNCTFWLQ 490

Db 413 HRALEASVARNRPDQGCISCMCNHTDMLY -SAKQYAAVVASDFFPRDPAS - - - - - 462

QY 491 GCHVH - - - - -CANDSLMGNFIHPDMDQSHNRCAPFAASRAISGCPITYSDSVGKHN 547

Db 463 - - - - -HVVHISVAYNTLFLGEFMQPDMDHSHLPAEYTGARALIGCPTIYSDKRGHN 520

QY 548 DLKLKLVLPDGSILRSEYVVALPTTRDCLFEDPLHNGETMLKIMNLKFPVGVGAFNQCQGG 607

Db 521 DLKRLVLPDGSVYLAOLPGRPTRRCLFSDPARQDASLTKIMNMKCAGVGVFGCQAG 580

QY 608 WCRETNRNQCFSQYSKRVTSKTPNKDIE - - - - -WHSGENDISIEGVKTEALYLYQ 657

Db 581 WCRVAKKTRIDEAPRGTLTGSVRAEDVAYIAQAAGTGDW -GGE - - - - - - -AVVYAH 629

QY 658 AKKILKSRSDDLQALDPFEFLITVSPVYKLIOTSHPAPICGLVNLNLNSGALQSDY 717

Db 630 AGELVLRPRGATLPVTLKRLLEYELFNHCPV -RAVAPGVSFAPITGLHFMNMGAAEECTV 688

QY 718 - - - - -DDDLSSVEIGVAGCGEMRVFASKRRACRIDEDGVGFYDQDQ -WVVYQVPP 769

Db 689 ETGEDGNAYVGLRVKRGCGRGATYCSRRPAKCSYSDADVETFTYDSDTGLVTADVPV 744

RESULT 10

043408 PRELIMINARY: PRM: 765 AA.

AC 043408;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE Putative lmbidation protein.

OS Brassica oleracea (Cauliflower).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Brassica.

OX NCBI_TaxID=3712;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. ALPHA BALOMA; TISSUE=ROOT TIPS;

RA Fujikura Y., Karszen C.K.;

RT "Cauliflower cDNA encoding a putative lmbidation protein.";

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

EMBL: X79330; CAAS5893.1; -

SR SEQUENCE 765 AA: 84084 MW: 733966030485958 CRC64;

Query Match	34.9%	Score 1473.5	DB 10	Length 765
Best Local Similarly	37.6%	Pred. No. 3.9e-11		
Matches 297	Conservative 155	Mismatches 253	Indels 85	Gaps 18

Qy 21 MSSPFAIDGSDPTVNGHFLSDVPENIVASRSPYTIIDKSPVSGCFVGFDASEDSRHV 80
:: :: : | : : : | : : : | : : :
Db 3 ITSNTSVQNDNLVVCGKITLTLPDNIILTPVAGASDS-----GAFICATFKQSKSLHV 57

```

QY 81 VSIKLDIQRMSLFERRKYMTTHMVRNNGDLESEFOIULK-----SDSGRRYVLF 134
Db 58 FPIGLLEBRMCSFRKRLMTMTOBKASGKDIPLERFOMLSKDEVCNDODAPTYVTVF 117
QY 135 LPIVEGPRPTSIQDDEDFVUDCVESGSSKRVVASFRRSMYLUHLAGDPPFLVKEAKIVR 194
Db 118 LPLLEGGRANVLOENKEKNEIEICLESDDKAVGTSQGHILYYVHAGNTPREVITQSVKAE 177
QY 195 THLGTFRLLEKTRPGIYDVKFGWCMTWDAFYLLVHNRQSVIEGVRHLVDGCGRPGLVLIIDG 254
Db 178 ROMOTFHNRREKKLPRSEVDFMGWCMTDAFYTDVTAEGVDGELSLSGCGTPRPFLLIIDG 237
QY 255 WOSIGH---DSDPTREKGMQVTAGBGMPCRLKFGQNTKFFRDYVNPVKATGRPAQOKMK 311
Db 238 WQOJENNEKSDNCNVQJ-----GAQFATRLVGIKMAKFOKN-DPKD---QVSGLK 285
QY 312 AFIDELKEFEFTVEHYVVMHNLGCVMGGLRPQVYGLP--EARYLOPRLPSGLQMTMEDLA 369
Db 286 SVYDNAK-QRINVAQOYVAMHNLAGVWGVGVPBRASGMHNYDSALATRYQSGVGLNGRDIY 344
QY 370 VDKIYVLRKGVLPPEKAEEMYEGJLAHLEKVGIDVKIDVILHLEMLCEDYGGRVLDKA 429
Db 345 MDSLAVHGLGIVNKRKYFNEFYNELHSYLASCGIDGVKDVQONIIETJLGAIGGRVSLTRS 404
QY 430 YKKAMPTYSINKHEFGNGCVIASMEHCNDPMLGLTPALISLGRVDGDFWCTDSSGDRNPNTFWL 489
Db 405 YHOLAEIASIARNEFKDNCISCMCHNTGVL-SAKOTAIVASDYYPRDAS----- 455
QY 490 QGCHMVHVCAN---SLWMGNFIHPDMQMFQSTHPCAFNAASRAISGGRPYVSDSVGKHN 546
Db 456 ---HTIHASAVAYNTLFLGGERMQDMDHMFHSNHTAETHAANAARVGCALATYVSDKRGKHN 512
QY 547 FDLKLKLVLPDGSILRSEYVALPTRDCLFEDPLHNGETMLKIMNLANKFTGVIGAFNCQGG 606
Db 513 FDLRLKLVLPDGSVLRARLRPERPRDCLFADPARDQSLTLKIMNMNKFITGIVFNCQGA 572
QY 607 GWCQETRRNQCFQSQYSKRYVSKTPKPI-----EMHSGENPISIEGVKFFALYLVQ 657
Db 573 GWCQDJKTKNRIHDTSPGTLGVLABEDADLLISEVAGODM--CGOS-----IYAYK 621
QY 658 AKKILILSKPSODDIALDPFEFLITVQSPYTKLQTSILHAPRIGLVMLNTSGAISQV-- 715
Db 622 SGELVRLPRKGSITVLTKVLEYELFHTISPL-KETIASISFAPRITGLDMFMSGAIQSMET 660
QY 716 ----DYDDDLSS-----VEIGKVGCGEMVRFAKPRACRIDIDGEDYGVKYDO 758
Db 681 NTVIDEKPELSSSSSVSENSRPTALLISLGVRCGCRFGAYSSQRPRLCAVDOGTEFENYDA 740
QY 759 D-QMAYVQVP 767
Db 741 EVGLTAVLNLPR 750

```

RESULT	11
Q94A08	
ID	Q94A08
AC	Q94A08:
DT	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	Putative imbibition protein.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eustroids II; Brassicales; Brassicaceae; Arabidopsi.
OX	NCBI_TaxID=3702;

RP
Sequence from N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Tortiumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carlinici P., Chen H., Cheuk R., Hashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neuman G., Kawai J., Kim C., Koestena E., Lam B.,

RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shim P., Southick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full length cDNA of gene T8H10.120/Atg57520 (GI:6706423)."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY050772; AAK92707.1; -
SQ SEQUENCE 773 AA; 85171 MW; 18BDF9E67952C801 CRC64;

Query Match 34.5%; Score 1459.5; DB 10; Length 773;
Best local similarity 36.6%; Pred. No. 5,5e-110;
Matches 294; Conservative 157; Mismatches 249; Indels 103; Gaps 18;

```
QY 21 MSSPFAIDGSDFTVNGHSLSDVNPENIVASPSPTYSIDKSPVSGCFVGFDPASPDSDRHV 80
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 3 ITSINISVQNDLVVQGTILTKIPDNILTP-----VTGNGFVSGSFTIGATFEQSKSLHV 57

QY 81 VSIGKLDIRFMSIFRRKVMWTTTHWGRNGDLESETOIYILEK-----SDSGRPVY 132
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 58 FPIGLVGLRMCCFRKKMMMTQRMSCGCKDIPLETQFMLESKDEVEGNGDAPVYVT 117

QY 133 FLPIYVGPRTSIQPGDDDFVDYCVSGSSKVVYDASFRSMLYLHAADDPRALYKEMKI 192
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 118 VFPLLEGQPRAYVLOGNEKEIEICFESGDKAVETSGQTHLVYVHAAGTNPREVRSQVKA 177

QY 193 VRTHLGTFRLLEKTPRGIVDFGMCWTDAFYLLTHVPOGVLEGVRLHYVDGCGPPGLVLD 252
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 178 VERHMOFHHREKKLPSFLDFMGWCMTDAFYDVTYTAEGVDEGLSKLSSEGTPPKFLIID 237

QY 253 DGMOSIGH---DSDPITKEGNOTVAGEQMPCKLLKFOENYKFRDYVNPKATGPRACQK 308
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 238 DGMQOIEKKERKDEMCVYRE-----GAQFATRLVGIKENAKFQ-----KSDQKD 280

QY 309 -----GMAFIDELGKEKTYEHVYVWHAALCYNGGLRPQVGLP--EARVQLPVLSTGLQ 362
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 281 TQVSGLSKSVYDNK--QRHNKQYVAMHALLAGYMGVGRPAASGMHYDSALAYPVQSGVL 339

QY 363 MTMEDLAVDKTLVLRKVGIVLPERKAEMEYEGLAHLEKVGIDGVKIDYIHLEMLCEDYGG 422
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 340 GNPDIYMDSLAVHGLGVLPKFKVFNFNELHSLASCGIDGVAVDQNIETLIGAGLIG 399

QY 423 RVDLAKAYKAMKRSINKHKGNGVIASMEHCNDFMELGTEAISLGRVGDDEFWCTDPSGD 482
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 400 RVSILTRSYQALASIAARNFTDNGCISCMCHNTDGLY-SAKQTAIVRASDFYRDPAS- 457

QY 483 PNGTFMLOGCHMYCAN---DSLMMGNFIHPDMDFOSTPCAFHNASRAISGPIYVS 539
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 458 -----HTHIASVAVNSLFLGEPMDMDMFHSLPRTAETVHAARAAGCAIYVS 507

QY 540 DSVGKHNFDLRLKVLPRDGSILRSEYVALPTRDCLFEDPLHNGETMLKIMLNKFTGVI 599
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 508 DKRGNNHFDLRLKVLPRDGSVLRKLPGRPTRDCLFADPARADGISLKLIMNMKFTGIVG 567

QY 600 AFNCGGAGCREFRRNOCFSQYSKRVTSKTNPKDI-----EMHSGENPISIEGVT 650
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 568 VFNCGAGMCKETKKNQIHDTSPGLTGSIRADADLISQVAGEDW--SGDS----- 617

QY 651 FALYLYOAKKLIILSKPSODDIALDPREFELITYSPVTKLIQTSLSHAPIGLVMLNTSG 710
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 618 -IYVAYRSGEVVRLPKGASIPLTILKYLELFIHSPL-KEITENISFAPIGLVDMFNSSG 675

QY 711 AIOGVD-----YDDLSL-----VEIGVAGCGEMERFASKKPRAC 745
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 676 AISISIDINHVDKNPEFFDEISASPALSDNSPTALVSVAGCGRFGAYSQRRLKC 735

QY 746 RIGDEYGEKYDOD-QMAYVQVP 767
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 736 AVESTETFDYDAEVLGVLTLNP 758
```

RESULT 12

Q9SCM1 PRELIMINARY; PRT; 773 AA.
ID Q9SCM1
AC Q9SCM1;

DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE Imbibition protein homolog.
GN T8H10.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansoorge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RP [2]
RA SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133248; CAB6109.1; -
SQ SEQUENCE 773 AA; 85143 MW; 0852F9E67952C8D3 CRC64;

Query Match 34.5%; Score 1458.5; DB 10; Length 773;
Best local similarity 36.6%; Pred. No. 6,6e-110;
Matches 294; Conservative 157; Mismatches 249; Indels 103; Gaps 18;

```
QY 21 MSSPFAIDGSDFTVNGHSLSDVNPENIVASPSPTYSIDKSPVSGCFVGFDPASPDSDRHV 80
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 3 ITSINISVQNDLVVQGTILTKIPDNILTP-----VTGNGFVSGSFTIGATFEQSKSLHV 57

QY 81 VSIGKLDIRFMSIFRRKVMWTTTHWGRNGDLESETOIYILEK-----SDSGRPVY 132
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 58 FPIGLVGLRMCCFRKKMMMTQRMSCGCKDIPLETQFMLESKDEVEGNGDAPVYVT 117

QY 133 FLPIYVGPRTSIQPGDDDFVDYCVSGSSKVVYDASFRSMLYLHAADDPRALYKEMKI 192
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 118 VFPLLEGQPRAYVLOGNEKEIEICFESGDKAVETSGQTHLVYVHAAGTNPREVRSQVKA 177

QY 193 VRTHLGTFRLLEKTPRGIVDFGMCWTDAFYLLTHVPOGVLEGVRLHYVDGCGPPGLVLD 252
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 178 VERHMOFHHREKKLPSFLDFMGWCMTDAFYDVTYTAEGVDEGLSKLSSEGTPPKFLIID 237

QY 253 DGMOSIGH---DSDPITKEGNOTVAGEQMPCKLLKFOENYKFRDYVNPKATGPRACQK 308
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 238 DGMQOIEKKERKDEMCVYRE-----GAQFATRLVGIKENAKFQ-----KSDQKD 280

QY 309 -----GMAFIDELGKEKTYEHVYVWHAALCYNGGLRPQVGLP--EARVQLPVLSTGLQ 362
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 281 TQVSGLSKSVYDNK--QRHNKQYVAMHALLAGYMGVGRPAASGMHYDSALAYPVQSGVL 339

QY 363 MTMEDLAVDKTLVLRKVGIVLPERKAEMEYEGLAHLEKVGIDGVKIDYIHLEMLCEDYGG 422
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 340 GNPDIYMDSLAVHGLGVLPKFKVFNFNELHSLASCGIDGVAVDQNIETLIGAGLIG 399

QY 423 RVDLAKAYKAMKRSINKHKGNGVIASMEHCNDFMELGTEAISLGRVGDDEFWCTDPSGD 482
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 400 RVSILTRSYQALASIAARNFTDNGCISCMCHNTDGLY-SAKQTAIVRASDFYRDPAS- 457

QY 483 PNGTFMLOGCHMYCAN---DSLMMGNFIHPDMDFOSTPCAFHNASRAISGPIYVS 539
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 458 -----HTHIASVAVNSLFLGEPMDMDMFHSLPRTAETVHAARAAGCAIYVS 507

QY 540 DSVGKHNFDLRLKVLPRDGSILRSEYVALPTRDCLFEDPLHNGETMLKIMLNKFTGVI 599
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 508 DKRGNNHFDLRLKVLPRDGSVLRKLPGRPTRDCLFADPARADGISLKLIMNMKFTGIVG 567

QY 600 AFNCGGAGCREFRRNOCFSQYSKRVTSKTNPKDI-----EMHSGENPISIEGVT 650
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 568 VFNCGAGMCKETKKNQIHDTSPGLTGSIRADADLISQVAGEDW--SGDS----- 617

QY 651 FALYLYOAKKLIILSKPSODDIALDPREFELITYSPVTKLIQTSLSHAPIGLVMLNTSG 710
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 618 -IYVAYRSGEVVRLPKGASIPLTILKYLELFIHSPL-KEITENISFAPIGLVDMFNSSG 675
```

OY 711 A10SVYD-----YDDLSS-----VEIGKGGCKEMRVFASKKPRAC 745
 DB 676 A1ESIDINHVTDKNPEFFGGEISSASPALSDNRSPTALVSVSRGCGRGVASSQAPLAC 735
 OY 746 RIDGEDVGFYDOD-QMNVVQVP 767
 DB 736 AVSETDETYDAEVLVTLNLP 758

RESULT 13
 O8RX87 PRELIMINARY: PRT: 844 AA.
 AC O8RX87;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE AT5g20250/ES024_140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayshtzki Y., Ishida J., Jones T., Kamuya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.W., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onders C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Soultani A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY090237; AAL90901.1; - 2AC8AB0EA43F8056 CRC64;
 SQ SEQUENCE 844 AA: 94215 MW: 2AC8AB0EA43F8056 CRC64;

Query Match 34.2%; Score 1446.5; DB 10; Length 844;
 Best Local Similarity 39.5%; Pred. No. 7.2e-109;
 Matches 305; Conservative 140; Mismatches 255; Indels 73; Gaps 18;

OY 20 DNSPPAIDSD--FTVNGHSFLSDVPENIVASPSPTYSIDKSPVSGCFVGDASEPDS 77
 DB 95 DMTIKPAVRISDGNLIKNTILTGVPDVIYTT---SASEAGPVE-GVFVAGVFNKEES 149
 OY 78 RHVYSGLKIDREMSIFRKVMWTHWGRNGDLESETOIILEKSDSG----- 128
 DB 150 KHTVPIGLTNRSMFCFRKTLMMMAORMGEMGRDIPYEQFLLVESNDGSHLESOGANG 209
 OY 129 -----RPVFLRPVIEGPFTSIQPGDDFVDVCESSSKVVDASFRLYLHAGDDPF 183
 DB 210 VECNQVYVYFELPLIEGSPSCQGNVNDVELCLESGDVTDRSRSTHSLYIHAGIDPF 269
 OY 184 ALVKEAMKIVRHLGTFRLLEKTPRGIVDKFGMCTWDAFYLVHPDQVIEGVRLHLDG 243
 DB 270 QITTDIAIRYKTLNLSFRQHEKKLPGIYDFGMCWDAFYQVTEGVAAGLKLAAAG 329
 OY 244 CPGGLVLIDGMSIGHDSRPIKEGMNQV-AGEQMP-----CRLLKFOQNKYFRQVNP 258
 DB 330 TPKFYIIDDGWSVEDEA-----TVEAGDEKKEPSIFRLTGIKENKFK----- 374
 OY 299 KATGPAGGKGMKAFIDELGKFEKTEHYVWHALCGYMGGLRPQVYLPPEARVIO-PVL 357
 DB 375 KKDPPVNGIKNIYKIKAKEKIG-----LKYVYVWHAITGYMGVNR---PGEYGSVMKTPNM 427
 OY 358 SPGLQMTMEDLAVDKIVLHKVGLVPEKAEEMEGLAHLEKVGIDGVKIDVTHLEMLC 417
 DB 428 SKGVENDPTWKTDVMTLQGLVSPKKYKFFYNELHSLYADAGVGVKADVCQVLETLIG 487
 OY 418 EDVGGVADLAKAYKAMKTSINKHFGKNGVYIAMEHCNDMFLEKTEISLGRGDDPFMT 477
 DB 488 GGLGHNVELTROHQAOLDSSVARNFPDNGCIACMSHTDLYCSKQAAYI-RASDDPYR 546

OY 478 DPSGDENGTFWLOGCHMVHCAN---DSLWGNFIHPDMDFOSTHPCAAFHASRAISCG 534
 DB 547 DPVS-----HTIHASVAVNSVFLGFERMQDMDMFHSHVHAEAHNASRAISCG 595
 OY 535 PIYSDSVGKHNFDLKKLVLPDGSILRSEYVALPRLDCLFEDPLHNGETMLKIMLNKF 594
 DB 596 PLVSDSPCKHNFELRLKLVLPDGSILRLRALPGRPTKCLFADPARDGVSLKIMWNKY 655
 OY 595 TGVIGAFNCOGGWCKRETRNOCFSQYSKRVYKSTKPNKDEMS-----GNPISIEGVKT 650
 DB 656 TGVIGVNCOGAAMSSSTERKNIFHQRTDLSIGRQV--HSISEASTDPTWNG--D 711
 OY 651 FALYQAKRLILSKPSODIALDPFELIVSPVTKLIQTSLHFAPIGLVNLMTSG 710
 DB 712 CAYSSOSRGLIYMPYVNSLPSVLKIREHIFTVSPISHLVD-CVSFAPIGLVNMNNSG 770
 OY 711 A10SVYDDDLSSVEIGVKGCGEMRVFASKKPRACRIDGEDVGFYDQDQV 763
 DB 771 A1EGLRYEAEKMKVMEVKGCGKFGSYSVKPRCVESNEIAFEYDSSGLV 823

RESULT 14
 O9M4M7 PRELIMINARY: PRT: 779 AA.
 AC O9M4M7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Putative seed imbibition protein.
 GN SIP.
 OS Persea americana (Avocado).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
 OX NCBI_TaxID=3435;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-CV. HASS; TISSUE-MESOCARP;
 RA Zamorano J.P., Evans A.D., Dopico B., Lowe A.L., Wilson I.D.,
 RA Merodio C., Grierson D.;
 RT "Isolation and characterization of cDNAs for mRNAs regulated during
 RT cold storage of avocado (Persea americana Mill.) fruit."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ131448; CAB77245.1; -
 SQ SEQUENCE 779 AA: 85368 MW: C3A8B43160316785 CRC64;

Query Match 33.5%; Score 1413.5; DB 10; Length 779;
 Best Local Similarity 36.7%; Pred. No. 3.1e-106;
 Matches 293; Conservative 144; Mismatches 268; Indels 93; Gaps 17;

OY 26 AIGSDFTVNGHSFLSDVPENIVASPSPTYSIDKSPVSGCFVGFDAEPDSRHVYSGK 85
 DB 8 SINDGNLVHNGKTIILGVPDNIYLPRTGDGL-----VACFCIGATNSESSESHVPMGT 62
 OY 86 LKDIREFSIKRVWMTTHWGRNGDLESETOIILEKSDSGRP-----YVFL 134
 DB 63 LEGIRFCCRFKLMWMTQMGKGVPLETOFMLETESKGAALIDDDEEAPTYIVF 122
 OY 135 LPVIEGPFTSIQPGDDFVDVCESSSKVVDASFRLYLHAGDDPFRLVKEAMKIVR 194
 DB 123 LPLEQFRAVLQGNESNQIELESIDGCAVRKNQWYLVYHMGAPRFPVINDAVKAVE 182
 OY 195 THLGTFLLEKTPRGIVDKFGMCTWDAFYLVHPDQVIEGVRLHLDGCGRPGLVLDG 254
 DB 183 KHLXSOHLEKKLIPSLDMFGMCTWDAFPTDYTDEVEEGSLSLSGGTPRRLIIDDG 242
 OY 255 WQISIGHSDPITKEGMNQVIA-GEOMPCRLKFOENYKFRDYVNPKATGPAGGKGMKAF 313
 DB 243 WQIGSDE--TDDNSCVVXEGAQFASRLTGIKENDK-----QKNGKSEHVPLGLV 293
 OY 314 IDELGKFEKTEHYVWHALCGYMGGLRPQVPLP--EAARYQPVLSPGLOMTMEDLAVD 371
 DB 294 VDAK-QHNHVKEVYVWHALAGYWGKVPRAAGMEHYDTLALAYPVOSPGVGNOPDIVMD 352

```

OY 372 KVLHKKVLLVPEKAEKMEYEGHIAHLEKVGIDGKIVIHLEMLCEDYGGGRDLAKAY 43
OY 373 SLVSGDGLVHPREKRYENYNELHAXLSCGVNGKVDQNTTEFLGAGHGERVSLTNSYI 412
Db 353
OY 432 KAMTKSINKHFHKGNGVIA SMEHCNDFMFLGTETALISLGRVGDDEWCTPSPGDPNGTFWLOG 491
OY 433 GALESIARNEFPDNCICMCHNTDSYI-SAKOJAVYRASDPFPRDPAS----- 461
Db 413
OY 442 CHNVH---CAADSLMNGEHIHPDDMFQSTHTPCAFAHAAASALISGPIYSDSVSKNED 548
OY 443 HTTHVSSVANNSSLFELGEMOPDMDFMSHLPAEYHGAARAVGCEPIYSDKRGHNFE 520
OY 549 LKTKVLVLDGSLISSEXYALPTPCLEFDPPLHNEETMLKINLNKFPVGICAFNCGCGW 608
Db 521 LKTKVLVLDGSLYLRALRPLRGPTROSLFYDPARDGVSLLKTIANNKKGICAGVGFPNCGAGW 580
OY 609 CRETRRNOCFSOYSKRYTSKTNPDIE-----WHSGENPISIEGVTFALTYLOAK 659
Db 581 CKTKTKTHIHDAAPOTLSGSIKRAHDEVEFINLAGQDM-NGE-----VIFFTYSG 629
OY 660 KLIISKPODDIALDPPEEFELITYSPYTKLIQTSLSHAPPLGLVNMINTSAIOSVD--- 716
Db 630 EVVRLPKCASIPVTLLEVLELXHCIPV-KEITNISIFAPGLGDMNSGAEVQFQVRM 688
OY 717 -----YDDDS-----VEIGVKGCGEMRFYFAKSKPRACRIDGEDV 752
Db 669 DSNNAPEPLFGYKASKSLSSLSNQSASAVLVLRVAGCGFAGYSSORPLKCTIVLDET 748
OY 753 GFKIDQ-DQMYVVOVQWMP 769
Db 749 EFNYDSVTGLVTLIIIPV 766

RESULT 15
O9LFZ7 PRELIMINARY; PRT; 1170 AA.
O9LFZ7
AC 09LFZ7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE F20N2.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RA SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Khan P., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F20N2 from chromosome
RT 1."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RA SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RA SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RA SEQUENCE FROM N.A.
RA Shinn P.P., Buehler E.E., Dunn P.P., Feng J.J., Kim C.C., Li Y.Y.,
RA Walker M.M., Altafi H.H., Araujo R.R., Conn L.L., Conway A.A.B.,
RA Gonzalez A.A., Hansen N.N.F., Hlizer L.L., Klemenetskaia I.I.,
RA Lenz C.C., Li J.J., Liu S.S., Luros S.S., Rowley D.D., Schwartz J.J.,

```

RA Toriumi M.M., Vysotskaia V.V., Yu G.G., Davis R.R.W.,
RA Federspiel N.N.A., Theologos A.A., Ecker J.J.R.;
RL Submitted (Dec-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Makharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologos A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC002328; AAF79504.1; -
DR InterPro: IPR004798; Cax.
DR TIGRFAMs: TIGR00378; Cax; 1.
SQ SEQUENCE 1170 AA; 129057 MW; 98B43A04E3F66D44 CRC64;

Query Match: 27.0%; Score 1140; DB 10; Length 1170;
Best Local Similarity 33.3%; Pred. No. 1,1e-83;
Matches 259; Conservative 135; Mismatches 208; Indels 176; Gaps 20;

QY	26	AIDSDPFWNHSLSIDVPEMIVAS--SPYTSIDKSEVUSVCGFEGFASDPDSHHVYIG	84
Db	8	SVTSDSLVVLGHVRLHGVPEMVLVTPASGNALID-----GAFIGVNDOTGSHRVSILG	61
QY	85	KLKIDIRFMSITRFKVVWVTHHVVGNNGDLSBEIYVILEKS-----DSGRPVYFLP	136
Db	62	KLIEDIRFVCVRFKLMMTQOMNGNGEILPECEIYTLIVENQSGDLGRODSSSYVFLP	121
QY	137	IVEGPFRTSIOPGDDDEVDVCSGSSKKVDASPRSMILYHAGDDPALVKEAMKIV---	193
Db	122	ILBEDFRAVLQGNANELEICLEGSDPTVQOFESHSILVFAAGSDPRPDVITKAVKYMFL	181
QY	194	RTHLGT-----PRLBEKTPRGIVDKFRCMGTWDAFLYVHPOGVLEBGVRLVD-----	241
Db	182	KSOELKTSLCRNFEM-----PDLIMNFQMCWTDAFYTNATKAVKQGLSENCDLTKPAL	235
QY	242	-----GGCPRGVLJLDDQSGISGHSDPTTEKGMQNTVAGEOMPRLKFOENKFRD	294
Db	236	ILCSLKAGGYPRKVIYIDDGQSGMGEYSVEFNADN-----AANFARLTHIKENHKFQ-	290
QY	295	VYNEKATGPRAGQOKMKA-----FIDELKEGFTEVHYVYVHVALCGYWGGLRPOV	345
Db	291	-----KDKGEGHRDPELSLGHVITDIDIKSN--NSLKVVYVYVHVALITGYWGGLRPGVS	340
QY	346	GLP--EARYIOPVLSPCGIOMTMBELADKILYHNVGLYPRPKAEEMEGJAHLEKVID	403
Db	341	GMEHYESVNAVYSSPGWSSSENGCLESITKMKGLGVNPEKVPSTYNDJHSLTASYVD	400
QY	404	GVRIDVILHLELCEDEYGRDLAKAYKAMKTSINKHEKNGCVITASMEHNCNPFMLETE	463
Db	401	GKVKDYQVNIILETDLAGHGGRYKLAKKYNHOLLEASISKNRPDNGIISCMSINTDGLV-SAK	459
QY	464	AISLGRGVDFMCTDPSGDDPNGTWLOGCMHNVHCAN--DSLWNGNFIHPDWMFOSTHR	520
Db	460	KTAIVIRASDWFMPDPS-----HTIHIAASYAATMLTDLGERFMQDWMFHSLLP	508
QY	521	CAAFHAASRAISGGRITVSDSGKHNFDLKLKVLVPGSLTSLREYVALPRDCLFEORPLH	580
Db	509	MAEIHAAARAAGCATIYSDKPGOHDPNLLRKVLVLRGSLTKRAKLFCGRPLRELY-----	563
QY	581	NGETMLKIVNLNKETGVIGAFNCGGGMWCRETRRNOCFSOYSKRVYTSKTPKDIEMHSGE	640
Db	564	-----LPKD-----	567
QY	641	NPISIEGVKTFALVLYOAKKLIILSPQDIDLADPREFELIVSPYTKLITQSLHFAPY	700
Db	568	-----TSPLVTLMPREYEVFLVYVVPVKKESDGS--KFAPV	599

Fri Apr 4 12:31:33 2003

us-09-675-208-5.rspt

Page 11

Oy 701 GLVNM LNTSGAIO SVDYDDLSS -VEIGVKCGGEMRVEAS-KKPRACRIDEDYGFK 755
|| : | : || : || : : : || : : || :
Db 600 GLMEKMSGGAIVSLRYDDEGCKFVVRMKLRGSGGLGVGYSSVRPRPSVTVDSDDV EYR 657

Search completed: April 4, 2003, 11:12:35
Job time : 48 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:09:06 ; Search time 48 Seconds

(without alignments)
1570.196 Million cell updates/sec

Title: US-09-675-208-5

Perfect score: 4225

Sequence: 1 MAPSFKNGSNVVSFDGLND.....QVWPIDSSGSGISVIEYLF 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1853.5	43.9	807	2 C65025	hypothetical prote
2	1482	35.1	757	2 S27622	Sid1 protein - bar
3	1473.5	34.9	765	2 S45033	probable imbibitio
4	1458.5	34.5	773	2 T46188	imbibition protein
5	1140	27.0	1170	2 C96599	protein F20N2.14 f
6	490.5	11.6	357	2 T09530	probable seed inh
7	471.5	11.2	371	2 T01717	hypothetical prote
8	365	8.6	649	2 D90496	hypothetical prote
9	121.5	2.9	538	2 S76481	hypothetical prote
10	115.5	2.7	902	2 T41051	beta transducin -
11	114	2.7	623	2 T42245	probable polypepti
12	113.5	2.7	632	2 S69702	hypothetical prote
13	105	2.5	730	1 T64118	1,4-alpha-glucan b
14	103	2.4	3944	2 T19997	hypothetical prote
15	102.5	2.4	544	2 G90524	hypothetical prote
16	102.5	2.4	564	1 HMI17	hemagglutinin prec
17	102.5	2.4	582	2 S74819	extracellular solu
18	102.5	2.4	1093	2 T51503	valine-tRNA ligase
19	102	2.4	679	2 G71615	phospholipase A2-1
20	101.5	2.4	478	2 S26222	mannose-1-phosphat
21	101.5	2.4	626	2 T42246	polypeptide N-acet
22	101	2.4	514	2 A36793	hypothetical prote
23	101	2.4	1195	2 T50226	hypothetical prote
24	101	2.4	1195	2 B96746	probable kinesin T
25	100.5	2.4	712	2 G02512	interleukin-1 rece
26	100	2.4	520	2 S27197	hydroxymethylgluta
27	100	2.4	1550	2 S60228	glutamate synthase
28	99	2.3	770	2 S76095	hypothetical prote
29	99	2.3	979	2 G90459	formate dehydrogen

30	99	2.3	1137	2 T18625	atrial natriuretic
31	98	2.3	306	2 D69753	vismycin phosphor
32	97.5	2.3	624	2 T42247	polypeptide N-acet
33	97	2.3	520	2 A25332	hydroxymethylgluta
34	97	2.3	520	2 S12736	hydroxymethylgluta
35	97	2.3	859	1 VCLJ22	env polypeptide
36	97	2.3	1002	2 T09438	toxR-activated gen
37	97	2.3	1013	2 B82276	genome polypeptin
38	97	2.3	3061	1 JN0545	hypothetical prote
39	96.5	2.3	444	2 S48408	DNA polymerase B1
40	96.5	2.3	1095	2 E84210	phospholipase C (E
41	96.5	2.3	1095	1 A31225	allantoinease (EC 3
42	96	2.3	460	1 S48489	hydroxymethylgluta
43	96	2.3	520	2 S45497	1,4-alpha-glucan b
44	96	2.3	737	2 C81724	AAA family ATPase
45	96	2.3	769	2 E90158	

ALIGNMENTS

RESULT 1	
C65025	hypothetical protein AT4g01970 (imported) - Arabidopsis thaliana
C:Species:	Arabidopsis thaliana (mouse-ear cress)
C:Date:	16-Feb-2001 #sequence_revision 16-Feb-2001 #texL_change 16-Feb-2001
C:Accession:	C65025
R:Anonymous:	The European Union Arabidopsis Genome Sequencing Consortium, The Cold SP
A:Title:	Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number:	A65001; MUID:20083488; PMID:10617198
A:Accession:	C65025
A:Status:	preliminary
A:Molecule type:	DNA
A:Residues:	1-807 <STO>
A:Cross-References:	GB:NC_001268; NID:97268581; PIDN:CAB80690.1; GSPDB:GN00140
A:Gene:	AT4g01970
A:Map position:	4
Query Match	43.9%; Score 1853.5; DB 2; Length 807;
Best Local Similarity	44.5%; Pred. No. 1.5e-138;
Matches	366; Conservative 146; Mismatches 257; Indels 53; Gaps 14;
QY	1 MAPSFKNGSNVVSFDGLND--SSPFAI-----DGSDFTVNGHSFLSDY 43
DB	1 MAPLHE-----SLSSINDVIESKPLFVPIITKPIIQPNFNLSEGLCAKDSPTPLFDV 53
QY	44 PENIVASP-SPYTSIDKSPSV-----GCFVGFDPASEPDSRHVVSIGKLDIREM 92
DB	54 PQNVFTPTPSSHSISIDAPLPIILRVQANAHKGGFLGFTIKESPSDGLTNSLGFEDREFL 113
QY	93 SIFRFKVMVTHVNGSGDLESETOIVLE--KDSGRRVYFLPIVGGPRTSIOPOD 150
DB	114 SIFRFKMMVSTAWIGKSGSDLOAETQWMLKIPEDS---VAIIPTIGAFRASTLPGE 170
QY	151 DDFVDVCGSSGSKVVDASFRSMVLYHAGDDPFAIVLEAMKIVRTPLGPRFLIEKTPPG 210
DB	171 KGNVLICAGSGSTKVKESSEFSKSTAYTHICDNPMLKKEAPSAIRVMMTFKLEEKLPK 230
QY	211 IYDKFGMCTMDAFYLVHPOGVIEGVRLVVDGCGPGLVLIDGQMSIGHDSDPTRKEGM 270
DB	231 IYDKFGMCTMDACVLYVDPATITWGVKKEEDGVCYKFIYIDGQMSINPDGDELKDME 290
QY	271 NOTVAGEQMPRLKQENYKFRDYVNPRAKATGPRAGQCKMAFIDELKEEFTVEHYVW 330
DB	291 NVTLGGEQMTARLTSEKCKKFRNFKESLSGSDVSGSMAAFTKDLRLRFRSLSDIYVW 350
QY	331 HALCGVWGLRQVPLPARYIQVPLSPGLQMTMDLAVDKIVLHKVGLVPEKKEEMV 390
DB	351 HALCGVWVREPTMDLAKVAPFELSPSLGATVMDLVVDDKVEAGSGLVHPSKHEFY 410
QY	391 EGIHAHLERKVGIDGVKIDVYIHLEMLCEDYGGKRVDLAKAYVYKAMTKSINKHFGKNGVIA 450

```

      : : : : | | : | | | | | : : : : | | | | : : : : | | | |
Db 411 DSHMSTIASVGYGAKIDVFOITLESLEEHGGVVELAKAYDGLTESMKNFNGTVIAS 470
451 MENCNDFMLGTFAISLIGVDDFWCTDPSGDNNGTFMLQGCAMVCANDSLMGNF IHP 510
471 MOOCNFFELATKOJISIGRVDDFWODPYGDDQGVYVLOGVMTHTCSINSTIMMGOMIQP 530
511 DMDMFQSTHPCAFHAASRAISGSPITYSDSVK--HNFDLKLVLPDGSILRSEYAL 568
531 DMDMFQSDHVCAYHAASRAICGGPYVLSLDHGLKASHNFDLKKLAFPDGTIRCVHYAL 590
569 PPRDCLFEDPLNGETMLKIMNLKFTGYGAFNCGGCGKCFETRNQCFSQSKRYTSK 628
591 PPRDSLFKNPLFDEKESILKIFNFNKGVGIVGTNCGAGASPEHRRKKGKCYTTVSGT 650
629 TNPDKLEWMSGEPISIEGVKTFPA---LYLVQAK-LIISKPSQDDIALDPEFELL 683
651 VHVSDLEW--DONPEAGSGVYITGDYLYVKQOSEELLENKSEAMKITLEFSADLLS 708
684 VSPVTKLIQTSLHFAPIGLVNLMTSGAIOQSDYDDLLSSVEIGVKCGEMRVASKRPR 743
709 FVPVTELVSSGVAFADLGLIMNENCYGVQDMKVTGD-NSIRVDVKGEGRFMAVSSAPV 767
744 ACRIDGEDVGFKKYDQD-MVYVQVPPRPIDSSSGGIVIEYLF 784
768 KCYLNDKEAFKWEETGKLSFFVWV--EESGGISHLSTFT 807

```

RESULT 2

S27762

Slp1 protein - barley

C:Species: Hordeum vulgare (barley)

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997

C:Accession: S27762

R:Heck, G.R.; Dorsett, C.; Ho, T.H.

submitted to the EMBL Data Library, February 1992

A:Description: Cloning and characterization of a gene, Slp1, associated with seed imbibition

A:Reference number: S27762

A:Accession: S27762

A:Molecule type: DNA

A:Residues: 1-757 <HEC>

A:Cross-references: EMBL:M77475; NID:g167099; PID:g167100

C:Genetics:

A:introns: 64/2; 151/1; 181/2; 197/3; 489/3; 557/2

Query Match 35.1%; Score 1482; DB 2; Length 757;

Best Local Similarity 39.4%; Pred. No. 4e-109;

Matches 306; Conservative 128; Mismatches 266; Indels 76; Gaps 18;

```

QY 28 DGSDFTVNGHSLSDVNPENIVASPSPTYSIDKSPVSGCFGDSSEPRHVSIGLKL 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11 DGR-LAVRGRITVLSGVDNVTAAHAAGALVD-----GAFVGATAAEAKSHHFTGTLR 64
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 88 DIRFMAIFRFKVMWTHVNGRNGDLESETOYILE-----KSDSGRYVFL 134
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 65 DCRFMCLFRFKLMMWMTQRMGTSGRDVPLRLEQFLLEVPRAAGDDSDSDGDESEVYLV 124
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 135 LPIVEGRFRTSIOPGDDDEVDVCEVSGSSKVVVDASFRSMYLLAHAGDPPALVKEAKIVR 194
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 125 LPRLBOFRVTLQGNODELOICIESGDKAVETEOGMNWNVYVYAGINPFEVITQAVKAVE 184
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 195 THLGFRLLLEKTRPGIIVDKFGMCTWDARFLVNVPOGVLEGVNHVLVDGCGPRGLVLIIDG 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 185 KHTQTFHNRREKKTVPFVDFWFGMCTWDARFYTDVTAAGVKGQSLSLAEGAPRFLIIDDG 244
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 255 WQSIGDS--DPTKGMNQTYAAGEOMPCRLKLFQENYKFRDYVNPKATGPRAGQKGMK 312
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 245 WQOIGSENDP---GV-AVDEGAQFASRLTGIIRNTKTFQSEHNEET-----PGLKR 293
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 313 FIDELGKEFKTEVYVNHALCGYMGSLRPVPGRL--EARVIOPLVPSGLQMTMEDLAV 370
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 LVDETKKE-HGVKSYVYVNHAMAGYMGVAFSAAGMEHYEPALATPVQDSRGVTONQPDIV 352

```

102(6) clm 34

```

QY 371 DKIVLRKVGSLVPEKAEEMEGSLAHLEKVGIDGVKIDVILHLEMLCEDYGGVNDLAKAV 430
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 353 DSLVSLGLGLVHRRVHREYDELHAYLAACGVGVKYVDQNIYETLGAGCHVALTRAV 412
431 YKAMTSINKHFKGNGVIAEMHCNDPMFLGTEAISLGRVGDQPMCTDPSGDNNGTFWLO 490
443 HRALEASVARNFPDNCISQCHNTMLY-SAKOTLVNRRSDPFYPRDPS-----462
491 GCMHVV---CANDSLMNGNFIHPDWMFOSTHPCAFHAASRAISGSPITYSDSVKKNF 547
463 --HTVHSSVAYNTLFLGERMOPDMDFHSLHPAEVHGAAARIAGGPIYVSDKPGHNH 520
548 DLKLKVLDPGSLIRSHRYALPTROCLFEDPLNGETMLKIMNLKFTGYGAFNCGG 607
521 DLLRKLVLPGSVLRAQLPGRPTROCLFSDPARDGA SLKIMMNNKAGVGVFNCGAG 580
608 WCRETRNRNOCFSQSKRVNYSKTNPKDIE-----MHSGENPISIEGVKTFALYLO 657
581 WCRVAKKTRIHDEAPGLTGSVRADEYEAIAQAAGTGM-GGE-----AVYTAHR 629
658 AKRLILSKPSQDDIALDPEFELLTVSPYTKLIQTSLHFAPIGLVNLMTSGAIOQSDY 717
630 AGEVLRLPRGATLPVTLKRLLEYELFHVCPY-RAVARGVSFAPIGLLHMFNAGAVECTV 688
718 ---DDILSSVEIGVKCGEMRVASKKPRACRIDGEDVGFKKYDQD-MVYVQVPP 769
689 ETGEDGNAVVGLRVGRGFRGAYCSRPAKSVSDASADVEFTYDGLVYADVPV 744

```

RESULT 3

S45033

probable imbibition protein - wild cabbage

C:Species: Brassica oleracea (wild cabbage)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997

C:Accession: S45033

R:Fujikura, Y.; Karsen, C.K.

submitted to the EMBL Data Library, May 1994

A:Description: Cauliflower cDNA encoding a putative imbibition protein.

A:Reference number: S45033

A:Accession: S45033

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-765 <PU>

A:Cross-references: EMBL:X79330; NID:g488786; PID:g488787

Query Match 34.9%; Score 1473.5; DB 2; Length 765;

Best Local Similarity 37.6%; Pred. No. 1.9e-108;

Matches 297; Conservative 155; Mismatches 253; Indels 85; Gaps 18;

```

QY 21 MSSPFAIDGSDFTVNGHSLSDVNPENIVASPSPTYSIDKSPVSGCFGDSSEPRHVS 80
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 ITNISVQNDNLVVOGKTILTKIPDNILLPVAGAGSDS-----GAFIGATPFQKQSLHV 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 81 VSLGKLKDTRFMSIFRFKVMWTHVNGRNGDLESETOYILE-----SDSGRYVFL 134
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 FPLIGVLEGRLEMCCEFRKLMWMTQRMGTSGRDVPLRLEQFLLEVPRAAGDDSDSDGDESEVYLV 117
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 135 LPIVEGRFRTSIOPGDDDEVDVCEVSGSSKVVVDASFRSMYLLAHAGDPPALVKEAKIVR 194
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 118 LPRLBOFRVTLQGNODELOICIESGDKAVETEOGMNWNVYVYAGINPFEVITQAVKAVE 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 195 THLGFRLLLEKTRPGIIVDKFGMCTWDARFLVNVPOGVLEGVNHVLVDGCGPRGLVLIIDG 254
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 ROMQGTGHNREKKTVPFVDFWFGMCTWDARFYTDVTAAGVKGQSLSLAEGAPRFLIIDDG 237
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 255 WQSIGH---DSDPTKGMNQTYAAGEOMPCRLKLFQENYKFRDYVNPKATGPRAGQKGMK 311
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 238 WQOIGSEKSNCLVDE-----GAOFATRLVGLIEENAFQKN-DPKDT---QVSGLK 285
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 312 AFIDELGKEFKTEVYVNHALCGYMGSLRPVPGRL--EARVIOPLVPSGLQMTMEDLA 369
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 286 SVYDNMK-QHNHVKQYVYVNHAMAGYMGVAFSAAGMEHYEPALATPVQDSRGVTONQPDIV 344

```



```

      8 SVYSDSLVYLGHVHLGVPENLVLTAPASGNALID-----GAFIGVTSDDTGGSHRVFSLG 61
      85 KLKIDIFMSIFERKVMWTTTHMVRNGDLESEFQIVILEKS-----DSGRFYELLP 136
      62 KLEDDLRFMCYERKMMWORMGTNKGKIPCEQFLIVEANGOSDLGROSSSYVEFLP 121
      137 IVEGPRFTSIQPGDDDFVDCVBSGSSKYVDASFRLMYLHAGDDPFLVKEAMKIV-- 193
      122 ILEGDRRAVLQGNANELELICESGDPVTDQFEGSHLVFEVAAGSDPEFDVTKAVKYMFL 181
      194 RTHLGT-----FRLLEKTPPGIVDKFGMCTDAFYLYVHPGVIEGVNHLVD----- 241
      182 KSQLKTSICLPNFRM-----PDLNMFQCTWDADFYTNTADVQKGLSENCDLTKPAL 235
      242 -----GCCPPLVLIIDGMOSIGHSDPITKEGNMOTVAGGOMCRLLKPOENKFRD 294
      236 ILCSLKAGCVTPKFEVLIIDGMOSVGDENISVEFNADN-----AANFANRLTHIKENHKFQ- 290
      295 YNPKATGPRAQCKMKA-----FIDELKGEFKYEHVYVMAHLCGYWGGLRPQVP 345
      291 -----KDGKEGHRVDDPSLSLGHVITDIKSN--NSLKYVYVMAHITGYMGVYKPGVS 340
      346 GLP--BARYIQPLSPGLQMTMEDLAVDKIVLHKVGIYPRPEKAEEMYLEGHAHLEKVGID 403
      341 GMEHYESKVAYPSPSPGVWSSENGCLESITTKGLGLVNPKEKYSFENDLHASYLASVGD 400
      404 GVKIDVTHLEMLECEBYGSRVDLAKAYYKAMTKSINKHFKNGVIVSMEHCHNFMFLGTE 463
      401 GKAVDYNILLETIGAGHGRVAKLAKKYHQALESISRNFNDGITISMSHNTTGLY-SAK 459
      464 AISLGRVGDFFWCTDPSGDPNGTFLQGCHMVHCAN--DSLWGNFTHPDMWPOSTHP 520
      460 KTAIVASDDFMRDPAS-----HTIHASVAAYNTLPLGFMQDMDMFHSLHP 508
      521 CAAFHAAARSAISGPIYVDSVGKHNFDLKLKLYLPDGSILKSEYALPFRDCLFEDPLH 560
      509 MAEYHAAARAVGCAIYVSDKPGQHDNFNLRLKVLRLDGSILKRLKGRPRRELIVY----- 563
      581 NGETMLKINMLNFTGVIAFANGCGGCMGCRFTRNOCFSQYSKRVSKTYPKDIEMHSGE 640
      564 -----LPRD----- 567
      641 NPISIEGVKTFALYLYOAKKLLIKSPSODIALDPFEFELLITVSEVTKLIQTSLHAPAI 700
      568 -----TSLPYTLMPREVEVFVYVKEPSDSS-KEAPV 599
      701 GLVNMINTGAIQSVDYDDLS--VEIGVKGCGEMKRVAS-KKPRACRIDGSDVGFK 755
      600 GLMEMFNSGGAIVSLKRYDDEGTKEFVVRMKLRGSLGVGVSVARPRSVTYVSDSDEYR 657

RESULT 6
T09530
probable seed inhibition protein - chickpea (fragment)
C:Species: Cicer arietinum (chickpea, garbanzo)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T09530
R:Cervantes, E.
submitted to the EMBL Data Library, February 1996
A:Reference number: 216718
A:Accession: T09530
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-357 <CER>
A:Cross-references: EMBL:X95875; NID:g1212811
A:Experimental source: germinating seed
C:Genetics:
A:Gene: sip

Query Match      11.6%: Score 490.5; DB 2: Length 357;
Best Local Similarity 32.5%: Pred. No. 6, 2e-31;
Matches 127; Conservative 55; Mismatches 154; Indels 55; Gaps 11;

```

```

      187 KEAMKIVRTHLGTFRLLLEKTPPGIVDKFGMCTWDAFYLTVHPQVIGEVNHLVDGSCP 246
      3 QQAVKVAEKHMQFHHREKKRVPSFLIDFMGCTWDAFYLTVAEGVEGLKSLSEGSTPP 62
      247 GLVLIIDGMOSIGHSDSPITKEGNMOTVAGEOMPRLK-----FQENYK--FRDIYNPK 299
      63 RFLIM-----TWGNRLKLVKOSKSGCVGEAGOLLGDMIRDA 101
      300 ATGPRAQOK-----GMAKAFIDELKGEFKYEHVYVMAHLCGYWGGLRPQVPGLPE--AR 351
      102 NSKRNQNDQEQIPLGLKHLVDGK--KHNVKDYVYVMAHLAGYGVKPRPAATGMEHDTAL 160
      352 VIQPVLSPLQMTMEDLAVDKIVLHKVGIYPRPEKAEEMYLEGHAHLEKVGIDGVKIDVH 411
      161 ATSAVTRSS--TWNQPDIVMSLAVHGLRLVHPRGFTSTNE--THAYLASGVDGKVDVQT 217
      412 LLEMLCEDYGGKRDVLAKEYKAMTKSINKHFKNGVIVASHEHCHNFMFLGTEAISGRVG 471
      218 LRPPLVLDIVVESRLHAAIIMRLRLPLVTLNNGCIACMCHNTDGLY--SAKOTAIVRAS 276
      472 DDFWCTDPSGDPNGTFLQGCHMVH--CANDSLWGNFTHPDMWPOSTHPCAAFHAA 528
      277 DDFYRHPDAS-----HTIHISVAAYNSLPLGCFEMQPDWDFHSLHRAEYHAAA 325
      529 RAISGPIYVDSVGKHNFDLKLKLYLPDGS 559
      326 RAIGGCOFMSVISQATITLIFLRSVLADGS 356

RESULT 7
T01717
hypothetical protein A_I6002N01.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
R:Schneel, P.; Magg1, L.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana I6002N01.
A:Reference number: 214407
A:Accession: T01717
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-371 <SCH>
A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191156
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 5/2; 25/3; 103/3; 126/1; 174/2; 240/1; 284/1; 295/1; 353/2; 349/2
A:Note: A_I6002N01.5

Query Match      11.2%: Score 471.5; DB 2: Length 371;
Best Local Similarity 26.9%: Pred. No. 2, 1e-29;
Matches 131; Conservative 71; Mismatches 112; Indels 173; Gaps 18;

```


A:Cross-references: SGD:S0002827; MIPS:YDR419w
A:Map position: 4R

Query Match 2.7%; Score 113.5; DB 2; Length 632;

Best Local Similarity 21.2%; Pred. No. 1.2; Mismatches 197; Indels 169; Gaps 23;

Matches 113; Conservative 55; Mismatches 197; Indels 169; Gaps 23;

```

QY 277 EQMPCLLK-----FOENYKFRDYVNPATGPRAGOKGKAFIDELKGEFTV 324
DB 39 EQMRGLSEDEPVVGVQNMSTIAVSAARKYGISRMDTIOELAKCSNLIPIHTAVFKKG 98
QY 325 EHVYVNHALCGYWGGLRQVPGLEPARYIQVLSQGLQMTMEDLAVDKIVLKGVLP-- 382
DB 99 EDFWQYHDCGSGW-----VODPAKQISVED-----HKVSLPEYR 132
QY 383 --PEKAEEMYDEGLNHLKLVGIDGVKIDYIHL--LEMLCEGYGGRVDAKAYKMTKS 437
DB 133 RSRRAALKIEFKSACDLVERASIDEVFLDGRICFNMMLMDNEELTGDLK--LKDALSN 189
QY 438 INKHFKNGVIASMEHCNDFMELGTEAI-SLGRVDDFWCTPSPGDPNGTFW----- 488
DB 190 IREAFIGWY-----DINSHLPLIEPKIKSLKFEQGVF--NPEGRLDTDMDDVILALG 241
QY 489 -----LOGCHVHCANDSLMNGNFIHPDW-----DMFQSTHPCAER 525
DB 242 SOVCKGIRDSIKDILGYTTSCLSTSKNCKLASNYKKPDAQTIYKNDCLDLFDGKFE 301
QY 526 AASRAISGPIYVSDSVGKHNFDLLKLVLP-----DGSILR 562
DB 302 ITSPWTLGVGL-----GKELIDVLD--LPHENSIKIHRETWPONAGOLKEFLDKKVO 352
QY 563 SEVYALPFRDCLFEDPLHNGETMLKIMNLN-----FTGVIGAFNCGGG-- 607
DB 353 SDY-----DSTSNIDPLKADLAELKFLSRGRYGLPSRPVVKSMNKNLKGSCMS 408
QY 608 -----WCRE-TRRNOCFSQ-YSKRVTSKT-----NPKDIEHSGENPISIEGVKT 650
DB 409 IYDCISMLEVFCAELTSRIQDLEQENKIVIPRTVASISLKTSEYEVRRKSGPAAVKGIN- 467
QY 651 FALVLYQAKLL--SKPSQDLDIALDPREFELITYSPYTKLQTSYLHAPICL 702
DB 468 -----FOSHELLKVGIKVYTDLDI-----KGNKSYPLTKLSMTITNPDIDL 511

```

RESULT 13

164118

1.4-alpha-glucan branching enzyme (EC 2.4.1.18) - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: 164118

R:Flaeschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Goeyne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: 164118

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-730 <TIGR>

A:Cross-references: GB:U32815; GB:L42023; NID:g1574818; PIDN:ANC23004.1; PID:g1574820; T

C:Superfamily: 1.4-alpha-glucan branching enzyme

C:Keywords: glycogen/starch biosynthesis; glycosyltransferase; hexosyltransferase

Query Match 2.5%; Score 105; DB 1; Length 730;

Best Local Similarity 17.5%; Pred. No. 7.3; Mismatches 203; Indels 272; Gaps 29;

Matches 117; Conservative 75; Mismatches 203; Indels 272; Gaps 29;

```

QY 68 VGDASEPDSRHVSTGKLDKIDRFNLSIFRKWMTTHWGRNG--DDESETQIYLK 124
DB 138 VNRFLMAPNARRSYVGD-----NYWGGRRHPRHRSKSGVWELFLPK 181

```

```

QY 125 SDSGREYVELLPIVGE-----PRTSIOQDDDFVQDCEVSGSKVYDASFRSLYHL 177
DB 182 ASLGGQLYKRELIDCHGNLKLKADPFASFSSQLRPTDASQVSAIPNVYEMTEARKKA----N 237
QY 178 ACDDPEALVKEAMKYRTHLGTFRLLLEETPPGIDYKFGCTWDAFYLTVPQGYEGVR 237
DB 238 QGNOP-----ISLYEVLGSMR-----RNLENRF-WLDYDQIADBLIPYVKEMGPT 282
QY 238 HL-----VDGGRPELVILIDGQWOSIGHSDPITKEGNNQYVAGQMCRLLKPE 288
DB 283 HLEPLSEFPDGSW-----GYOPLGLYS-PTSFGSPENR-----RLK-- 323
QY 289 NYKFRDYVNPATGPRAGOKGKAFIDELKGEFTVENYVNHALCGYWGGLRQVPGLP 348
DB 324 -----RAHEAGINVLIDWVPGHFPDTH----- 346
QY 349 EARYIOPVLSQGLQMTMEDLAVDKIVLKGVLPPEKAEEMTEGLNAN----- 396
DB 347 -----GL-----VAEDGTALY-----EHEDPREGYHODMNTLITVYGRNE 381
QY 397 -----LEKVIDGVKIDYIHL-----EMICEGYGGRVLD-AKAYKA 433
DB 382 VKNFLSSNALYWERFGVGIKRYDAVASMYIYNDYSAEBEMIPNOTGRENLEALEFLKH 441
QY 434 MKRSINKHFKNGVIASMEHCNDFMELGTEAISLGRVDDFWCTPSPGDPNGTFWLOGCH 493
DB 442 TMWKI--HSEMGALISAEESTSPAGV--THPSENGSLGNF-----KMMNG-- 484
QY 494 VYHCANDSLMNGNFIHPDWDFOSTHPCAFAHNASIAISGPIYVSDSVGKHNFDLLKLV 553
DB 485 -----WMDTL-----AYMKLIDYRQYHNHMKMFGVYQY 515
QY 554 VLPDGSILSEVYALPFRDCLFEDPLN-----GFTMKIMNLNFTGVIGAF-- 601
DB 516 -----SENFVPLS--HDEVHGYSLGKMPKQWQFAVLRAYVYGIMGYPGK 563
QY 602 -----NCQ-----GGGWCRETRR-----NOCFSQYSKRVTSKT 629
DB 564 KLFMNGEFAQGRNMYEESLDWFLDENIGGSMHNGVGLVKVLDNQIYOKNRPLELDN 623
QY 630 NPKDIEW 636
DB 624 SPEGPDW 630

```

RESULT 14

T19997

hypothetical protein C47D12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19997

R:Gajadasy, S. submitted to the EMBL Data Library, March 1996

A:Reference number: Z19209

A:Accession: T19997

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3944 <WIL>

A:Cross-references: EMBL:Z69902; PIDN:CA93765.1; GSPDB:GN00020; CESP:C47D12.1

A:Experimental source: clone C47D12

C:Genetics:

A:Gene: CESP:C47D12.1

A:Map position: 2

A:Introns: 46/3; 308/3; 408/2; 449/3; 820/3; 946/3; 1069/1; 1634/3; 1737/2; 2009/2; 2

Query Match 2.4%; Score 103; DB 2; Length 3944;

Best Local Similarity 18.5%; Pred. No. 1.4e+02; Mismatches 136; Conservative 90; Mismatches 230; Indels 280; Gaps 34;

Matches 136; Conservative 90; Mismatches 230; Indels 280; Gaps 34;

```

QY 80 VVSTGKLDKIDRFNLSIFRKWMTTHWGRNGDDESETQIYLKSDSGRPVYFLPIVE 139
DB 2931 VVSYGHVKTIDQAMNIVR-----EIEESTDIRI----- 2957

```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:01:51 ; Search time 76 Seconds
(without alignments)
1374.585 Million cell updates/sec

Title: US-09-675-208-5
Perfect score: 4225
Sequence: 1 MAPSPKNGSNVSPFDGLND.....QVMPPISSGGISVIEYLF 784

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A.Geneseq_101002:*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4225	100.0	784	19	AAW53570
2	4225	100.0	784	20	AAV17417
3	2996.5	70.9	777	20	AAV32074
4	2945	69.7	783	23	ABB93664
5	2860.5	67.7	783	20	AAV32073
6	2854.5	67.6	781	19	AAW57887
7	2854.5	67.6	781	20	AAV30143
8	2854.5	67.6	781	22	AAW49400
9	2852	67.5	758	21	AAV70978
10	2834	67.1	780	22	AAW98659

11	2767	65.5	799	19	AAW57886
12	2767	65.5	799	20	AAW30142
13	2529	59.9	587	19	AAW57888
14	2529	59.9	587	20	AAV30144
15	2364	56.0	572	20	AAV32075
16	1853.5	43.9	807	23	ABB92806
17	1490	35.3	751	21	AAV70980
18	1470	34.8	763	21	AAV70977
19	1432.5	33.9	756	21	AAV70975
20	1398.5	33.1	841	21	AAV70981
21	1394	33.0	750	20	AAV17418
22	1384	32.8	770	21	AAV70976
23	940	22.2	508	21	AAV70974
24	873	20.7	265	20	AAV32072
25	788	18.7	271	19	AAW57889
26	788	18.7	271	20	AAV30145
27	389.5	9.2	229	21	AAW28628
28	263.5	6.2	193	21	AAW28629
29	181	4.3	30	19	AAW53567
30	181	4.3	30	20	AAV17422
31	177.5	4.2	120	21	AAW28630
32	153	3.6	63	23	ABP35067
33	111	2.6	1174	13	AAW25825
34	109	2.6	1390	18	AAW21636
35	107.5	2.5	1584	20	AAV33727
36	107	2.5	487	19	AAW64182
37	107	2.5	487	23	AAU95497
38	107	2.5	488	19	AAW64159
39	107	2.5	488	23	AAU95482
40	107	2.5	1174	11	AAU08257
41	107	2.5	1174	17	AAW89494
42	107	2.5	1174	18	AAW09043
43	106.5	2.5	905	22	AAW20088
44	106	2.5	1080	22	AAW00803
45	106	2.5	1372	22	ABG14534

ALIGNMENTS

RESULT 1
AAW53570
ID AAW53570 standard; Protein: 784 AA.
XX
AC AAW53570;
DT 06-JUL-1998 (first entry)
XX
DE Cucurbit raffinose synthase.
XX
KW Cucurbit; raffinose synthase; sucrose; galactinol.
XX
OS Cucumis sativus.
XX
PN JP10084973-A.
XX
PD 07-APR-1998.
XX
PF 28-APR-1997; 97JP-0111124.
XX
PR 26-JUL-1996; 96JP-0198079.
XX
PR 26-APR-1996; 96JP-0107682.
XX
PA (AJIN) AJINOMOTO KK.
XX
DR WPI: 1998-264858/24.
XX
DR N-PSDB: AAV22250.
XX
PT Raffinose synthase gene - useful for preparation of raffinose in
XX transformed plant
XX
PS Claim 3; Pages 17-20; 26pp; Japanese.
XX

Broad bean raffin
Amino acid sequenc
Japanese artichoke
Amino acid sequenc
Raphesed raffinose
Herbicideally activ
Wheat raffinose sy
Rice raffinose syn
Corn raffinose syn
Wheat raffinose sy
Soybean raffinose
Rice raffinose syn
Corn raffinose syn
Amino acid sequenc
Arabidopsis thalia
Arabidopsis thalia
Cucurbit raffinose
Cucurbit raffinose
Arabidopsis thalia
Human ORF4040 prot
Novel toxin expres
Grapevine leafroll
Photographus lulin
Lettuce resistance
Lettuce pest resis
Lettuce resistance
Lettuce pest resis
B. thuringiensis to
B. t. toxin 81A2.
Bacillus thuringie
Rat CARD-6 protein
Human bone marrow
Novel human diagno

Qy	481	GDPMGTFLQCC	HNVCANDSLMNGNFTIHPDMFMQSTHPCAFHAASRAISGPIYSD	5400
Db	481	GDPMGTFLQCC	HNVCANDSLMNGNFTIHPDMFMQSTHPCAFHAASRAISGPIYSD	5400
Qy	541	SVGKHNFPLK	LIVLPQGSILRSFYVALPTRDCLFEEDPLHNGETMLKINMLNKPFTGVGA	6000
Db	541	SVGKHNFPLK	LIVLPQGSILRSFYVALPTRDCLFEEDPLHNGETMLKINMLNKPFTGVGA	6000
Qy	601	FNCGGCGC	RETRRNOCFSQYSKRVTSKTPNKDIEMHSGENPISIEGVTFFALTYOAKK	6600
Db	601	FNCGGCGC	RETRRNOCFSQYSKRVTSKTPNKDIEMHSGENPISIEGVTFFALTYOAKK	6600
Qy	661	LILSKP	SODDIALDPEFEBELITVSPYTKLIQTSILHAPAGLVNMLTSGAISOVDYDD	7200
Db	661	LILSKP	SODDIALDPEFEBELITVSPYTKLIQTSILHAPAGLVNMLTSGAISOVDYDD	7200
Qy	721	LSSVEIGV	KCGCEMRVPASKKPRACRIDGEDVGRKYQDDQNVVVQVPWPIDSSSGISVT	7800
Db	721	LSSVEIGV	KCGCEMRVPASKKPRACRIDGEDVGRKYQDDQNVVVQVPWPIDSSSGISVT	7800
Qy	781	EYLF	784	
Db	781	EYLF	784	

CC	AAV32074	standard; Protein: 777 AA.
XX	AAV32074	
AC	AAV32074;	
XX		
DT	17-JAN-2000	(first entry)
XX		
DE	Mustard raffinose synthase.	
XX		
KM	Raffinose synthase; mustard; transgenic plant.	
XX		
OS	Brassica juncea.	
FH		
FT	Key	Location/Qualifiers
FT	Misc-difference 210	
XX	/note= "encoded by ACR"	
PN	EP953643-A2.	
PD		
XX	03-NOV-1999.	
PF		
XX	27-APR-1999;	99EP-0107430.
PR	30-APR-1998;	98JP-0120550.
PR	30-APR-1998;	98JP-0120551.
PR	04-DEC-1998;	98JP-0345590.
PR	10-DEC-1998;	98JP-0351246.
PA	(SUMO) SUMITOMO CHEM CO LTD.	
PI	Matanabe E, Oeda K;	
XX		
DR	WPI: 1999-593144/51.	
DR	N-PSDB; AAV20209.	
XX		
PT	New sense and antisense genes, useful for altering the level of	
PT	raffinose in food plants -	
XX		
PS	Claim 26; Page 29-31; 55pp: English.	
XX		
CC	This sequence represents mustard raffinose synthase, a protein	
CC	that can bind a D-galactosyl group through an alpha(1-6) bond to the	
CC	hydroxy group attached to the carbon atom at the 6-position of the	
CC	D-glucose residue in a sucrose molecule to form raffinose. cDNA	
CC	(see AAV20209) encoding the enzyme was isolated from mustard	
CC	(Brassica juncea) leaf cDNA by PCR. Probes or primers generated from	

CC plant raffinose synthase genes (see AA20207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.

Query Match	70.9%	Score 2996.5	DB 20	Length 777
Best Local Similarity	70.0%	Pred. No. 2.3e-274		
Matches 555; Conservative	99	Mismatches 114	Indels 25	Gaps 10

QY	1	MAPSFKNGSGNVNSFD----	GLNDKSSP-FAIDGSDFTVNGHSLSDVPEINVASPSXY-	54
QY	1	MAPSFKNGSGNVNSFD----	GLNDKSSP-FAIDGSDFTVNGHSLSDVPEINVASPSXY-	54
Db	1	MAP-----PSYKSDAAVNNGIDLSGKPLFRLEGGDILLANHHVLTDPVNVVTYASPYL		54
QY	55	TSIDKSPV--SVGCVGFDA--SEPSRHRVVSIGKILKDIRFMSIFPFKXWMTTHWGRGG		11
Db	55	ADKOEPPDASAGSTIGTNLDGEPFRSRIVASIGKIRDIRFMSIFPFKXWMTTHWGRGG		11
QY	112	DLESETOJLILEKSDSGRPYVFLPRIVEGPRPTSIOQPDDEDFVDCVSSGSKVVDASPR		17
Db	115	DIENETQJLILENSGSRPYVLLPRLEBGSFRSSPQGEDDDVAVCBSGSGQYTGSEFR		17
QY	172	SMULYHAQDDPFRALYKEMKIVRTHLGTFRLLLEKTPRGYIDKFGKCTWDAFYLTVNRQG		23
Db	175	OVVYVYHABDDPRKYLKDAKMYVRVMMNFKLLLEKTPRGYIDKFGKCTWDAFYLTVNRQG		23
QY	232	VIEGVRHLYVDGCGPRGVLIDDDGMSIOGSDPRTKEBMONQVAGEQMRCLKPOENYK		29
Db	235	VHKGVKCLYDGGCPRGVLIDDDGMSIOGSDGIDVEGMSCTVAGEQMRCLKPOENYK		29
QY	292	FRDYNPRATGPRAGOKMKAFIDELAKEFRTVEHVYVMAHLCGYMGLRQVPLPEAR		35
Db	295	FRDYVSPK----DKNEVGMKAFRDLKEEFSTVDYIYVMAHLCGYMGLRQVPLPEAR		35
QY	352	VIQPLVSLGLOMTMDLAVNDKIVLHKVGLVPRPEKAEEMYEGHNAHLEKVGIDGVKIDVYH		41
Db	351	IYRPLSLGKRLTMDLAVNDKIVDTGIDGIFVSPDMANEEYEGHSHLQVNGVIGDGVKIDVYH		41
QY	412	LLEMCEDEYGGKRVDAKAYVRKMTKSIKKHFKGANGVISMENCHDNFMELGTAEALSIRVG		47
Db	411	ILEMCEKXGGKRVDAKAYFALYTSVAKNHEDGKNVISMENCHDNFMELGTAEALSIRVG		47
QY	472	DDFWCTDPSGDPNGTFWLIQGCCHVHCANDSLMGNFIHPDMDFOSTHPCAFAHNASPAI		53
Db	471	DDFWCTDPSGDPNGTFWLIQGCCHVHCANVSLMGNFIHPDMDFOSTHPCAFAHNASPAI		53
QY	532	SGGPIYVSDVSGKHNFDLLKLVLPDGSILRSEXYALPTROCLPEPDRLHNEBTMLKINL		59
Db	531	SGGPIYSDCVQOHFDLLKLVLPDGSILRCEHIALPTROCLPEPDRLHNDGTMLKINL		59
QY	592	NKFTGVIGAFNCGGMCRETRRNOCFOSQYSAKRYTSKTPKDIEMHSGENPISIEGVXTF		65
Db	591	NKYUTIIIGAFNCGGMCRETRRNOCFOSQCVNTLTATNPKDVEMNSGNPNISVENBEF		65
QY	652	ALYLYQAKKLLSKRSQDLIALDPFEFELITVSVPYTKLQTSLHFAPIGLVNLMLNTSGA		71
Db	651	ALFISQSKRLVYISGNDELLETLEPFKEELITVSVPVYIEGSSVQFAPIGLVNLMLNTSGA		71
QY	712	IQSVYDDDLSSVEIGVKGCGEMRFAASKPRACRIDEDGVFKYDDQOQMYVCPWPID		77
Db	711	IRSLVYHBE--SVEIGVAGAGEFRYVASKRASCRIIDEVYDFGCI--EBSMNVQVPM----		76
QY	772	SSSGGIVSIEVYLE 784		
Db	765	SAPEGLSSIKYEF 777		

RESULT 4

raffinose in food plants -
Claim 25; Page 22-24; 55pp; English.

CC This sequence represents sugabead raffinose synthase, a protein
CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC hydroxy group attached to the carbon atom at the 6-position of the
CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
CC (see AA20208) encoding the enzyme was isolated from sugabead cv.
CC hamling leaf cDNA by PCR. Probes or primers generated from plant
CC raffinose synthase genes (see AA20207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.

SQ Sequence 783 AA;

Query match 67.78; Score 2860.5; DB 20; Length 783;

Best Local Similarity 66.6%; Pred. No. 2e-261;
 Matches 530; Conservative 109; Mismatches 132; Indels 25; Gaps 12

Qy	1	MABEFKNGS----	NNVSFOGLNMSSP-FAIDOSDFTVNGHSLSDPVENIYASBPRT	55
Db	1	MABFSKSNSTCDEYVNHDCN--	-TPIISLEESNNVGHVILSOVPONI-----T	51
Qy	56	SIDSPPSVGCFVGFDASEPDSRHVVSIGRLKIDIRFMSIFRFKMWTTHVGRNGDLES	115	
Db	52	AISKMGFD-GLFVFDAPEPARHVRVSVGOLKGIPIFMSIFRFKMWTTHVTSNGRDLH	110	
Qy	116	ETQIVILEKSDS--GRPYVLLPIVBEPPRTSTOQCD-DDPVDCVBSGSSKVVADSFRS	172	
Db	111	ETQILILDKSDEGAGRPYVILPIEBEPFRASLOPGSVDDYVDVDCVBSGSTKVVAGDSFRA	170	
Qy	173	MLUYHAGDDPFAIYKEMAKIVRTHLGFTRLLKEETPGSIDYDKFGCMQWDAFFYLVHPQGV	232	
Db	171	VLYIRAGDPRPKLKDPMKEVQALHGFRLKLLDOKTPPGIYDKGCMQWDAFYLVKRYGV	230	
Qy	233	IEGVRLHYDGGCPGGLVLDIDGWSIGCHSDPIT-KEGMOYVAGEOMPRLKLFDEBNK	291	
Db	231	WEGYKGLVENVPGLVLDIDGWSIGCHDDPIIDOGIKRNTSAGEOMPRLIKYBENFK	290	
Qy	292	FRDYNPNKATPPR-AGQGMKAFIDELKGEFKYEHVYVNHALOGYGGRLPQVPGLEPA	350	
Db	291	FRDKSPKINCHEDHPMKMRAPYRDLKKEFKYEHVYVNHAFGLYGGVPRVNPGLPEA	350	
Qy	351	RVIOPLVSPGLQMTMEDLAVDKIVLHKVGLVPRPEKAEEMTEGLNHLKVGIDGVKIDVI	410	
Db	351	QVYVPRKLSPGLEMTMEDLAVDKIVNNGIGLQVOPKAOELKEGLSHLENGIDGVKVDVI	410	
Qy	411	HLLEMLCEDYGGRYLDKAVYKAMTKSINHPKNGNGVIASMEHONDMPLGTETLSIGRV	470	
Db	411	HLLEMAEDYGRYELKATYTKATTESYRKHFKONGVIASMEQONDMPLGTETLCIGRV	470	
Qy	471	GDDEMCSTDPSGDPNGTFMLOGCHVYMCANDSLWNGNFIHPDMDMFOTHPBCAFAHSA	530	
Db	471	GDDEMPDTPSGDINGTYWLOGCHVYMCAYNSLMWGNFIHPDMDMFOTHPBCAFAHSA	530	
Qy	531	ISGGPIYVSDYSGKHNDLKLVLPRGSLIRSEXYALALPRDCLFEDPLNGETMLKIYN	590	
Db	531	ISGGPIYVSDVYGNHNPIPLKRLVLAGSGSLRCEYHALPDKCLFEDPLHDGTMLKIYN	590	
Qy	591	LNKEFTGVIGAFNCGGCGMCFETRNNQCFQSYKSKVTSKTPKDIEMHSGENPISIEGVK	650	
Db	591	LNKYNGLVIGVENCGGGGMRSRKNLCEFEYSKRSICKTSBKVDWEMNGHKKPPKICVEC	650	
Qy	651	FALYLYQAKKILSKPSCODDIALDPEPELITYVSPVTKLIQTSLHAPRIGLVNMLTSG	710	
Db	651	FAMFTYERKKILISQLSDBTJESLIDPDELYIVVSPMTILPWSIAPRIGLVNMLNAG	710	

```
Oy      711 AIOSSVYDDLLS--VEIGVGCGCEMVFASKKPRACRIODEGFGFKYDODOMVVVOVPW    766
       ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      711 AVKSILDISEDEMDKKMQVQINGCAGBEMMYSSSKPKRACRVNGEDNEFEY-EESMKVGYTN    765
Oy      769 PIDSSGGISVIETLF   784
       :::|::|::|::|::|::|::|::|
Db      770 -NNHSGGTIVLEYLF   783
```

RESULT 6

ID	AAW
----	-----

57887 standard; Protein; 781 AA.

AC AAW57887;

DT 23-SEP-1998 (first entry)

DE Soybean raffinose synthetase

KW Raffinose synthetase; metabolism modification; food additive;

2000

[illegible]

XX
2
2
2
2
2

[illegible][illegible]

XX
XX

2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
96
98
100

XX 1000-334670/20
DD

DR N-PSDB; AAV40801.

PT New nucleic acid

PT beneficial effects on gastrointestinal flora
XY

PS Claim 1; Page 31-34; 44pp; English
xy

This sequence represents the soybean raffinose synthetase of the invention. The raffinose synthetase is capable of producing raffinose by combining a D-galactosyl group through an alpha (1-6) bond with a hydroxyl group attached to the carbon atom at position 6 of a D-glucose residue in a sucrose molecule. The DNA can be used to modify metabolism of a host organism by introducing into the host organism or its cell so that the content of the raffinose family oligosaccharides in the host organism or cell is changed. Raffinose oligosaccharides are useful as food additives with beneficial effects on the gastrointestinal flora.

SQ Sequence 781 AA;

Query Match	67.68; Score 2854.5; DB 19; Length 781;
-------------	---

Best Local Similarity 65.9%; Pred. No. 7.2e-261;
Matches 525; Conservative 123; Mismatches 120; Indels 29; Gaps 11

0y	1	MAPSRKNGSVNVPD-----GLNDMSPPAI-----DGDGFYNNHSPFLSDYDPAENVASPRSYTS	56
0y	1	1	
Db	1	MAPSI-----SKTVELNSLGLVNGNLPLESTLEGSNLFNGLANGPLFLTEVPENITVTPSP----	53
0y	57	ID-KSPVS-----VGCYFVGFDASPDSDRHVYSIGLKLDI RFMSI FRKRYVMYTHHWGRN	109
Db	54	IDAKSKNNEDDDVYGCYFVGFADEPRSRHVASIGLKLGI FMSI FRKRYVMYTHHWGSI	113
0y	110	GGDDESENOIYLEKSDS-GRPVYFLLPYVGPERTSIQPDQDFVYVCESSGSKVDA	168
Db	114	GHELEHOMMLIRKNDQGLRPFLVPIIDASFRASIQPDGLDYVYVCEMSGSTRWGSS	173

QY	169	SPRSALTYLHAGDDPFAIVKEMAKIVRPTLIGFFRLLEEKTPRGIVDKRGWGTWDAFEYTVH	229
Db	174	SFGSCLYHVHGHDPQLQLREKATKYVRNHLGFFKLLEKTAPEVITIDKRGWGTWDAFEYTVH	233
QY	229	PGQVTEGVRLHLDVGGCCPRGLVLDGCMOSIGHSDPIT - KEGMOTVAGEOMPCRLLFQ	287
Db	234	PSGVMEGKVLVEGGCPRGMVLDIDGCMOAIChDDPITIDGCMKRTSAGSQMPCRLVLE	293
QY	288	ENYKFRDQVNNKATGPRRAGQGMKAFIDELKGEKTEHNYVMHALLCGYNGSLRPOVGL	347
Db	294	ENYKFRQICSGDGS - - - - EKGMGAFVADLKEQFRSVQYVHVALDCGYNGKPRKPGM	348
QY	348	PEARVITQVLSPGLQMTMEDLAVNDKIVLHKGLVPRPEKAEEMVGLHAHLEKVIIDGKI	407
Db	349	POAKVVPFKLSNGKLITMKDLAVNDKIYNSGVGLVPRPHALHLYEGLSRLSAGIDGVK	408
QY	408	DYIHLLEMLCEDYGRVDLAAYTKAKTKSINKAFKNGVYASHEGNDPMEFLGTAISL	467
Db	409	DYIHLLEMLSEYGGRYVELAAAYKALASVKKHFKGVYASHEGNDPFLGTEAL	468
QY	468	GNVGDDEFMCTPSPGDPNPTFMFLQCCCHVHCANDSLMNGNFIHPMDMFQSTHPCAPHAA	527
Db	469	GNVGDDEFMCTPSPGDPNPTFMFLQCCCHVHCAYNSLMNGNFIHPMDMFQSTHPCAPHAA	528
QY	528	SRAISGGIYIVSDSYGKINFDLKKVLDPGSLIRSEYIALPTDCLFEDPLHNGETMLK	587
Db	529	SRAISGGVYIVSDCYGKINFEKLSLALPDGTILRCOHYALPTDCLFEDPLHDKCTMLK	588
QY	588	IWNLKKFPGVYIGAFNGCGGCGCRFTRNRCCSOYSKRTYSTKNDKDIEMHSGENPISIEG	647
Db	589	IWNLKKYGVGLERNGCGGCMPTTRNKKASSEFSQYVITCLASPDLEMSNGKSPICIKG	648
QY	648	VKTEFALYLYQAKKILSKSPDLDIALDPFEFELLITVSPVTKLIQTSLHAPRIGLVMLN	707
Db	708	TGSAQSVADYDDLLSVYIGVKGCGEMKVPFASKPRACRIDGEDVGRKYQDODMNVVQVP	767
QY	709	TGGALQSEFPNHNLDYVAKIGVRCGEMKVPFASKEPVCCKLDGVVVKPDY - EDKMLRVQVP	767
QY	768	WPIDSSGGSIVIEYLF 784	
Db	768	WP - - - SAKSLMVEPLF 781	
RESULT 7			
AAAY30143			
ID	AAAY30143 standard; Protein: 781 AA.		
XX	AAAY30143;		
XX	AC		
XX	26-OCT-1999 (first entry)		
XX	Amino acid sequence of a raffinose synthase protein.		
XX	Raffinose synthase; plant; sucrose; raffinose.		
OS	Glycine max.		
XX	JP11215984-A.		
XX	10-AUG-1999.		
XX	12-DEC-1997; 97JP-0342899.		
XX	28-NOV-1997; 97JP-0329006.		
XX	18-DEC-1996; 96JP-0338673.		
PA	(SUMO) SUMITOMO CHEM CO LTD.		
XX	WPI; 1999-51112/43.		
XX	N-PSDB; AA210002.		

PT New raffinose synthase gene - is prepared from a plant material
 PS
 XX
 XX Claim 8; Page 25-27; 40pp; Japanese.
 CC
 CC The present sequence represents a raffinose synthase protein. The
 CC sequence is isolated from plant material. The protein forms raffinose
 CC by complexing alpha(1 to 6)-D-galactosyl hydroxyl group of the 6C
 CC of D-glucose residue in sucrose molecules.

Query Match	67.6%	Score 2854.5	DB 20	Length 781
Best Local Similarity	65.9%	Pred. No. 7.2e-261		
Matches 555	Conservative 123	Mismatches 120	Indels 29	Gaps 11

OY	1	MAPSRKNGSUNVSPD--GLNDMSPAI--DGSDFTYNGHSFLSDVPENVASPSYTS	56
Db	1	MAPSI-----SKTVELNSGLVNGMLPLSLTLESGNFLANGHPLTEVPENIYTPSP---	53
OY	57	ID-KSPVS-----VGFVGFDASEPDSRHVSIGLKLDIREMSIFRFYVMTTHWYGRN	109
Db	54	IDAKSSKNNEDDDVYGVGFVGHADPEPRSRYHVASIGLRGKIFMSIFPFKVMWTHHWGSN	113
OY	110	GGDLESENOIYLKESOS--GRVYFLLPIYVGPRTSLQGGDDOFVNVYCSGSSKYDA	168
Db	114	GHELHEHIOMLLDKNDOLGRPFVLLPIYLQASFRASLOPDDIYVYVCHESGSTRVCGS	173
OY	169	SFRSMALYHAGDDPEFALVKEEMAKIVRTHLGTFRLLEKTEPRGIYDKRGMCWDAFYLVH	228
Db	174	SFGSLYVHGHDPYQILIREATKYVVRNHLGFKLLEKTAAPVILIDRKGMCTWDAFYLVH	233
OY	229	FOGYEGVRLHYDGGCGPRGLVLDGQWOSIGHDSPTT--KEGNOIYVAGEOMPCRLKFO	287
Db	234	PSGWEGYKGLVEBGCPCPMVLLDGOALCHDDPDTIDGOCMRTSAGOMPCRLVLE	293
OY	288	ENYKRDVYVNRKAGPRAGOKMKAFLDELKGEKTEVHYVNHALCGYWGSLRPOVPL	347
Db	294	ENYKTRQYCSKRD-----EKGMKAFYRDLKEDQRSVQYVNHALCGYWGVRPKPGM	348
OY	348	PEARVIOPLVSPGLQMTMEDLAVDKIYLHKYGLVPRPKAEEMYGHLNLEKGIQDVKI	407
Db	349	POAKVYVFKRLSNGKLITMKDLAVDKIYVNGGLVPRPLNHLVLEGLHSRLSAGISGVK	408
OY	408	DVILHLEMLCEDYGRVDLAKYVYKAMTKSLNKHFKNNGYIASMEHCNDFMLCTEASL	467
Db	409	DVILHLEMLESEYGRVELAAYYKALTAASYKHFKNGVYIASMEHCNDFMLCTEAL	468
OY	468	GRVGDEPFCPSDDPNCTFMLQCSHWYHCANDSLMNGNFIHPMDMFOSTHPCAFNHA	527
Db	469	GRVGDEPFCPSDDPNCTIYLOCSHWYHCAYNSLMNGNFIQPMDFOFSTHPCAEFNHA	528
OY	528	SRAISGCPHYVSDSVGKHNPLKLVLPDSILIRSEYVALPTRDCLFEDPLNGETMLK	587
Db	529	SRAISGCPHYVSDCVGKHNFELKSLALPDOTILRQHYVALPTRDCLFEDPLNHDKTMK	588
OY	588	IWNLNKFLPYGVGAENCGGCMCRETRNRQCSOYSKRYVSKTNPKDLEHMSGENPISIEG	647
Db	589	IWNLNKRYVYGLTELCGCGGCMCPYTRRNKSASESESOYVTECLASPODLEWMSGSPICIG	648
OY	648	VKTFALYLYOAKKILSLKSPSODDIALDPFEFELITYSPYTKLIQTSLHAPFGLVMNL	707
Db	649	MMNVRAVYLYFKRHKLKLMKASKLELVSLFPTELLITYSPYIVLSKLLIQAPFGLVMNL	708
OY	708	TSGAIOQVYDDDLSSVEIGVKGCGEMKRVASAKKPRACRILDGEDYGRKYDODOMVYVQ	767
Db	709	TGGAIOQWEPFNHIDVYKIGVKGCGEMKRVASEKRPVCSKLDGVVVKDY--EDKMLRVQV	767
OY	768	WPIDSSSGGISVIEYLF 784	
Db	768	WP-----SASKLSMVEFLF 781	

RESULT 8

AAB49400
ID AAB49400 standard; Protein; 781 AA.
XX
AC AAB49400;
XX
DT 07-MAR-2001 (first entry)
XX
DE Soybean raffinose synthase.
XX
KW Plant promoter; transgenic plant; desired property.
XX
OS Glycine max.
XX
PN EPI048733-A2.
XX
PD 02-NOV-2000.
XX
PE 27-APR-2000; 2000EP-0108962.
XX
PR 30-APR-1999; 99JP-0124527.
PR 01-SEP-1999; 99JP-0247211.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Ishige F, Watanabe E, Oeda K;
XX
DR MPI: 2001-104537/12.
DR N-PSDB; AAC89523.
XX
PT New soybean plant promoters useful for generating transgenic plants
PT with desired properties -
XX
PS Example 6; Page 24-27; 36pp; English.
XX
CC The present invention provides novel plant promoters which can be used in
CC the production of transgenic plants which express genes with desired
CC properties.
XX
SQ Sequence 781 AA;

Query Match 67.6%; Score 2854.5; DB 22; Length 781;
Best Local Similarity 65.9%; Pred. No. 7.2e-261;
Matches 525; Conservative 123; Mismatches 120; Indels 29; Gaps 11;

OY 1 MAPSKNGSNVYSPD--GLNDMSPPAI--DGSDFTVNGHSFLSDVPENIVASPSPTS 56
DB 1 MAPSI-----SKVELNSRGVLVNGNLPSTLLEGSNPLANGHFLREVPENIIVTSP--- 53
OY 57 ID-KSPVS-----VGEFVGFDASEPDSRHVVSIGKLDIRFMSIFRFKVMWTTHWVGRN 109
DB 54 IDAKSSKNEDDDVVGCFVGFHADERSRHVASLGKLRGIRFMSJFRFKVMWTTHWVGSN 113
OY 110 GDDLSESRQIVYLEKSDS-GRPYVELLPVEGPRFRTSIPGDDDFVDCVEGSSKVVYDA 168
DB 114 GHELEHETQMMLDLKNDQGRPFVLILPILQASFRASLPGLDDVYVDCMESGSTRVCGS 173
OY 169 SFRSMYLTHAGDPPALVYKAMKIVRTHLGTFRLLLEKPRPGIVDFGCTDAFLTVH 228
DB 174 SFGSCLYVNGIDPQQLREATKVVYRMHIGTKLEKTAPEYIIDKFGCTDAFLTVH 233
OY 229 PGQVLEGVHNLVDGCGPRGLVILIDGWSIGHSDPIT-KEGMNQTVAEQMPCRLLRQ 287
DB 234 PGVWEGVGLVEGGCPRPMLIDGQWALCHDEDPITDQEGMKRISAGEQPCRLVKLE 293
OY 288 EYKFRDYNPKATPRAGQKGMKAFIDELKGEFTEVYVWMLCGVWGLRQVPGU 347
DB 294 EYKFRQYCSGDS-----EKGMGAFLVRLKQDFRVEGVYVWMLCGVWGLRQVPGM 348
OY 348 PARVYIQPLVSLGLQWTEMDLAVDKVLHKVGLVPRPEKAEEMEYEGHLAEKVGIDGVI 407
DB 349 PAKAVTPRLSLNGKLTMDLAVDKVSNQVGLVPRPHLHLLEGLHSLKESGIDGVAV 408
OY 408 DIVHLEMLCEDYGGRVDLAKAVYKAMTKSINKHFKGNCVSIASMEHCNDPMLGTETALISL 467

DB 409 DIVHLEMLSEEGYGGRVDELAKAVYKALTAASYKKHFKGNCVSIASMEHCNDPMLGTETAL 468
OY 468 GRVGDDEFMCTDPSGDPNGTFMLQGGHMYHCANDSLMKGNFIHPMDMOMOSTHPCAAFIHA 527
DB 469 GRVGDDEFMCTDPSGDPNGTYWLOGCHMYHCAVNSLMGNFIQPDMDMOMOSTHPCAAFIHA 528
OY 528 SRAISGPRIVSDSVGKHNFDLKKVLVDPGSLIRSEYALPTPRCLFEDPLHNGETMLK 587
DB 529 SRAISGPRIVSDSVGKHNFDLKKVLVDPGSLIRSEYALPTPRCLFEDPLHNGETMLK 588
OY 588 IWNLNKFTGVIGAFNCGGCGMCETRRNQCFSQYSKRVYTSKTPNPKDIEHSGENPISIEG 647
DB 589 IWNLNKFTGVIGAFNCGGCGMCETRRNQCFSQYSKRVYTSKTPNPKDIEHSGENPISIEG 648
OY 648 VKTFAIYLYQAAKLLKSRPSQDLALDPFEELITVSRYTTLQSLHFAPIGLVNMNLN 707
DB 649 MNVFAVYLFKDKHLKLMKASEKLEVLSEPPFELLTVSPVILSKRLIOFAPIGLVNMNLN 708
OY 708 TSGAIOGVYDDDLSSVEIGVKGCGEMRYFASKKPRACRIDGEDYGFKYDODQVVOVVP 767
DB 709 TCGAIOGMEFDNHIDVVKIGVRCGEMKVFASEKFSCKLDGVVVKFDY-EDKMLRVQVP 767
OY 768 WPIDSSSGGISVIEYLF 784
DB 768 WPIDSSSGGISVIEYLF 781

RESULT 9
AAY70978
ID AAY70978 standard; Protein; 758 AA.
XX
AC AAY70978;
XX
DT 09-AUG-2000 (first entry)
XX
DE Soybean raffinose synthase from clone sfl1.pk125.d4.
XX
KW Soybean: raffinose synthase; raffinose saccharide;
KW clone sfl1.pk125.d4; nutritional; soy protein.
XX
OS Glycine max.
OS WO20024915-A2.
XX
PN 04-MAY-2000.
XX
PD 22-OCT-1999; 99WO-US24923.
XX
PR 23-OCT-1998; 98US-0105451.
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
PI Allen SM, Hiltz WD;
XX
DR MPI: 2000-350754/30.
DR N-PSDB; AAD00335.
XX
PT Nucleic acids and encoded proteins involved in the biosynthesis of
PT raffinose, useful for producing soybean seeds with a reduced raffinose
PT content and therefore improved nutritional quality -
XX
PS Claim 2; Page 47-49; 58pp; English.
XX
CC The present sequence is a raffinose synthase from
CC clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
CC library still. Raffinose synthase is involved in the biosynthesis
CC of raffinose and higher homologues in the raffinose saccharide family
CC from sucrose. The present sequence is useful for reducing the raffinose
CC saccharide content of soybean seeds which improves the nutritional
CC quality of the soy protein products derived from them.
XX
SQ Sequence 758 AA;

Query Match	67.5%;	Score 2852;	DB 21;	Length 758;
Best Local Similarity	67.5%;	Pred. No. 1.2e-260;		
Matches 530;	Conservative 94;	Mismatches 133;	Indels 28;	Gaps 9

QY	1	MASTRKNGSNVSPFDGDKNDKSSFPALDGDSPFYWNGHSLSDVPENIYVASPSPTYSIDKS	60
Db	1	MGPSKSKASAK-----SGYTKHMKGPSCLSNSTLKYNGVILLISQVKNVTLTPCTYDT----	52
QY	61	PVSVCPEFAPSEDSRHHVYSIGLKDIRFMSIFERKVMWTTTHWVRNGDLESEFOIV	120
Db	53	HTGCGFLGFHATSPKSKSHVAPLDQKNISPTSIFFRKVMWTTTLMTMSNGRDELETQFL	111
QY	121	ILKSDSRPPVFLPLPYEGFPRTSIQGDDEVDVDCVESSSSKVVYDASFRSMYLTHAGD	180
Db	112	MLQ-----SHPVLEPLPLQPPFRASLPDPSDDNVAVCESGSSHTVTSASPDYVYLHAGD	167
QY	181	DPFALVRFAKIVTRFLTLFLLLEEKTPPGIVDFKGCNTDAFLTYHPGVGLEGVHLV	240
Db	168	NPTFLYKRAMNVYAHNLSFLLLEEKTYPGVNDKFGMCTNDAPFLTYIHPGVEGVKGLV	227
QY	241	DGCGPPLVLIDDGMSQISGHDSDPTTBEGANQIVAGEOMPCLRLKEOENYKFRDYVNPKA	300
Db	228	DGCGPPLVLIDDGMSQISGHDSDPTTBEGANQIVAGEOMPCLRLSEDEENYKFRSY-----	281
QY	301	TGPRAGOKMAFIDELKGEFKTYEHYVWMLLCGYNGSLRPQVPGLEAPRVLPVLSPG	360
Db	282	---KKG-KGLGFPVRELKEEFGSVETIYVWMLLCGYNGVAPRGVAGAAEAVERKFLTEG	337
QY	361	LQMTMEDLAVDKIYLHKVGLVPRPEKAEMTEGLHAHLEKVGIDGVKIDVTHLLEMLCEDY	420
Db	338	LKGTMEDLAVDKIYVNGGVVPRPELVGEMEGTLHAHLESAGIDGVKIDVTHLLEMWCEKY	397
QY	421	GGRVDLAKAYYKAMTKSINKHFKNGVYASMEHNDNMPLEGTALISLGRVDDPFWCTDPS	480
Db	398	GGRVDMARAYTKALTYARKHFKNGVYASMEHNDNMLGTETALISLGRVDDPFWCTDPS	457
QY	481	GDPNGTFWLQSCHHVHCANDSLMANGNIHPDMDFOSTHPCAAFHAAASRAISGGPIYSD	540
Db	458	GDPNGTFWLQSCHHVHCAYNSLMMGNFIHPDMDFOSTHPCAAFHAAASRAISGGPIYSD	517
QY	541	SVGKHNPDLKKLVLPDGSILRSFYVALPTPDDCLFEDPRLNGEMLMKINMLNKFYVICA	600
Db	518	TVGNINPELLKTLPLPDSILRCHYVALPTPDDCLFADPLHDGKMLKINMINKYTYGLGV	577
QY	601	FNCGGGCMCFETRRNOCFSQYSKSVTSTKTNPKDILEMHSGEPISTIEGVKFFALYIYQAK	660
Db	578	FNCGGGCMCFETRRNOCFAEESHVSTKTNPKDILEMHSGEPISTIEBQVLFASFYSQAK	637
QY	661	LILSKPSODL-DIALDPEEFELTVSPYTKLIQTSLHPARIGLVNMLNTSGAIOQSDYDD	719
Db	638	LILSKPSODSEISILEPNEFLITVSPYTVLPKGSYVFAPIGLVNMMLNTGGAVOQSLAFDE	697
QY	720	DLSSEIVKSCGEMRFASKPKPRACRIDGEDVFXKDDOQMVVQVPMPLDSSGSISV	779
Db	698	GQNLVEVGLRGTGEMRYASASEKPTCTCIDGKENDFER-BGSMAVIQVPM-----GSSKLSIF	753
QY	780	LEYLF 784	
Db	754	VOYVF 758	
RESULT 10			
AAB98659			
ID	AAB98659 standard; protein: 780 AA.		
XX	AAB98659;		
AC			
XX			
DT	17-AUG-2001 (first entry)		
XX	Soybean protein: SEQ ID 1.		
DE			
XX			
FW	Mutant: muteln; raffinose synthase; raffinose oligosaccharide reducti		

KM	plant; soybean.
XX	
OS	Glycine max.
XX	
PN	JP200107873-A.
XX	
PD	27-MAR-2001.
XX	
PF	03-JUL-2000; 2000JP-0200571.
XX	
PR	09-JUL-1999; 99JP-0196036.
PA	(SUMO) SUMITOMO CHEM CO LTD.
XX	
DR	WPI: 2001-313373/33.
XX	N-FSDB; AAH27438.
PT	Novel mutant protein of raffinose synthase is useful for reducing the
PR	raffinose oligosaccharide content in a plant body -
PS	Disclosure; page 18-20; 30pp; Japanese.
CC	The present invention relates to a mutant protein of raffinose synthase
CC	in which at least one aromatic amino acid present at the position of
CC	about 1-7 amino acids from the N-terminus is deleted or replaced. The
CC	mutant protein can be used for reducing the raffinose oligosaccharide
CC	content in a plant body. The present protein from soybean, was used in
xx	the present invention.
SQ	Sequence 780 AA:
Query Match	67.1%; Score 2834; DB 22; Length 780;
Best Local Similarity	66.0%; Pred. No. 6.3e-259;
Matches 526; Conservative 121; Mismatches 120; Indels 30; Gaps 12;	
QY	1 MAPSRKNGSGNNVSD--GLNDMSSPFAI-DGSPFTYNHGHSFLSDVENIVASSPPTS 56
Db	1 MAPSI-----SKTVELNSGVLVGNPLSLTLEGSNPLANGHPFLETVEENIIIVPSP--- 53
QY	57 ID-KSPVS-----VGCFEFGFDASEPDSRKHVYSIGLKDIREFSIFRKVMVTTHWGRN 109
Db	54 IDAKSSKNNEDDDVVGCFEGFIADDEPRSHNVASLOKLGIKMSIFRRKVMVTTHTWGSN 113
QY	110 GDDESEFTQVILEKSDS-GREYVELLPIVEGPFFTSIQPGDDEVDVCVESGSKVVDA 168
Db	114 GHETHEHQOMLLDKNDQLGREPVILRLPLQASFRAISLQPGLDIDVDVDCMESGSTRVSGS 173
QY	169 SFRSMVLVHAGDGDPALVKEAMKIYRTHLGPRLLIEEKTPRGTVPKPGMCMTDAFYLVH 228
Db	174 SGSGCLYHAVGHDPQLQLREATKYVRMLGTFTKLEEETAPRYIVDKFCMCTIDAFYLVH 233
QY	229 PGGVTEGVKHLVDGGCPGLVLLIDGWOSTIGHDSPIT-KESGMOTVAGEOMPRLKFQ 287
Db	234 PGGVEWGKGLVEGSCPGMVILLIDGMOAICHDEDPRIIDQEMKRRTSAGEMPRLVLE 293
QY	288 EMYKRDVYNKRAIPRRAGQGKMAFIDELKGFEFTVEHYVWMHALCGYMGGLRQVPGL 347
Db	294 ENYKRFROYCSGDS----EKGMGAFFVKDLKEQFRSVQVYV-HALCGYMGGVKRVKPGM 347
QY	348 PEARYITQVLSFGLTOMTMEDLAVIDLTLYHKGYLVPEAEEMEYBELNAHLKVGIDGYKI 407
Db	348 PQAKVVFTRKLSGLTLTKMDLAVDKITVSNGVGLVPRPHLAILLYBSLRSLASAGIDGVKV 407
QY	408 DVIHLLLEMLCEDYGGRVDLAKAYYKAAMTKSTINKHFKMGCVIASMEHCDFMLGTIAISL 467
Db	408 DVIHLLLEMLESEYGGRVELAKAYYKALTASYKKHFKKGAVIASMEHCNDFLLGTIEATAL 467
QY	468 GRVGDDEFMCTPPSGPNGPMFLGCHMHVNCANDSLIMKCNFLHPDDMFQOSHPCAAFPAA 527
Db	468 GRVGDDEFMCTDPSGPNNGTYWLOGCHMHVCAVNSLMKCNFLTQPDMDFOSTHPCAEPFAA 527
QY	528 SRAISGGIYVSDVSGKHNFDLKLVLPPDSGISLSEYYVALPTRCLEDPEPLIHNETMIK 587


```

PF 12-DEC-1997; 97JP-0342899.
XX
XX 28-NOV-1997; 97JP-0329006.
PR 18-DEC-1996; 96JP-0338673.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
DR WPI; 1999-51112/43.
DR N-PSDB; AA210001.
XX
XX New raffinose synthase gene - is prepared from a plant material
XX
XX Claim 5; Page 19-21; 40pp; Japanese.
XX
XX The present sequence represents a raffinose synthase protein. The
XX sequence is isolated from plant material of broad beans. The
XX protein forms raffinose by complexing alpha(1 to 6)- D-galactosyl
XX hydroxyl group of the 6c of D-glucose residue in sucrose molecules.
XX
XX Sequence 799 AA:
SQ
Query Match 65.5%; Score 2767; DB 20; Length 799;
Best Local Similarity 64.0%; Pred. No. 1.5e-252;
Matches 508; Conservative 118; Mismatches 140; Indels 28; Gaps 8;
QY 11 NVVSEFDGLNDMSSP-FAT---DGSDFTVNGHSFLSDVPENIVASPSPTYSI-----D 58
DB 14 DVSTIDINGNSPLSTILDSRDLANGHPRLTQVPMTITTTTATSSFLNLSNKD 73
QY 59 KSP-----VSVCFCVGDASPPDSRHVYSIGKLDIRFMSFFRRKVMWTTIMVGRNG 111
DB 74 TTPNNNNMMLLOOGCFVGNSTEPKSHHVPGLGKLGKIFMSIFRRVWMTTHVGTNGQ 133
QY 112 DLESETQIVILEKSDS-GRPVYFLPTVEGPFRTSIOPGDDDFVDVCVSGSSKVVDAFP 170
DB 134 ELQHERQMLILDKNDLGRPVYLLPILFLENTFRTSLQPGLDNHLIGSVESGTHVGSST 193
QY 171 RSMLYIHAGDDPRPALVKEAMKIVRHLGTFRLLEKTPRGIVDKFCGCTWDAFYLTVHPQ 230
DB 194 KACLYTHLNDPSTLKEAVKVIOTOLGTFKLEEKTAISIIDKFGMCWDAFYLVKHPK 253
QY 231 GYIEGVRHIVDGGCPGLVILDDGQSIQHDSPRIKBCGNQVAGEOMPRLCKROENY 290
DB 254 GWMEGKSLTDGGCPFGVYIIDDGQSIQHDDEDDSGNRFSAEGOMCRVLKTEENS 313
QY 291 KFRDYVNPATGPRAGQOKMAKFIIDELKGEFTVEHVYVWMAHLCGVGGLRPOVLPRA 350
DB 314 KREYENPE-----NGGKKGILGCVRLDKBEFGSVESVYVWMAHLCGVGGRVPHGMKRA 369
QY 351 RYIQLVLSGLQMTMEDLAVDKIVLHKVGLVPRKEAEMVEGLHAHLEKVGIDGVKIDVI 410
DB 370 RYVVPKVSOGKMTMEDLAVDKIVENGVLVPRDFAHMEFDGSHLESIGIDGVADVAI 429
QY 411 HLEEMCEGYGGVNDLAKKYYKAMTKSINKHFGNGVILSMENCNPFMEGLPAISILGR 470
DB 430 HLEELISEEYGGVRLARAYKALTSYKHKRGNGVILSMENCNPFLLGDAISLGR 489
QY 471 GDDFWCTDPSGDPNGTFTWLQGGCHMVHACANDSLMGNFIPDMDFOSTPCAFHAASRA 530
DB 490 GDDFWCTDPSGDPNGTFTWLQGGCHMVHACANDSLMGNFIPDMDFOSTPCAFHAASRA 549
QY 531 ISGGPIYVSDSVGKHNFDLLKLVLPDGSILNSEYALPTRDCLFEDPLHNGETMLKINN 590
DB 550 ISGGPIYVSDSVGKHNFDLLKLVLPDGSILNSEYALPTRDCLFEDPLHNGETMLKINN 609
QY 591 LKFGVIGAFNGCGGCGMRETRRNCQFSQYKRVTSKTNPRDIEMHSENPISIGVVT 650
DB 610 LKFGVIGAFNGCGGCGMRETRRNCQFSQYKRVTSKTNPRDIEMHSENPISIGVVT 669
QY 651 FALYLYQAKLLKSPSODLIALDPFEFELITVSPVTKLIQTSLHPAIGLVNMLNTSG 710
DB 670 FAVYFFEKKKLLMKMSDKLVKSLPEFSEFLMTVSPYKVFSSKRFIOFAPRIGLVNMLNSG 729

```

```

QY 711 AIGSVYDDDLSSVEIGKCGCEMRVFAKKRPACRIDGEDVGFKKYDQDMVVQVWPPI 770
DB 730 AIGSLEFDDNASTLVKIGKCGCEMRVFAKKRPACRIDGEDVGFKKYDQDMVVQVWPPI 787
QY 771 DSSSGISVLEYLF 784
DB 788 -SSSTLSLVQFLF 799
RESULT 13
AAW57888
ID AAW57888 standard; Protein; 587 AA.
XX
XX AAW57888;
XX
XX 23-SEP-1998 (first entry)
XX
XX Japanese artichoke raffinose synthetase.
XX
XX Raffinose synthetase; metabolism modification; food additive;
XX gastrointestinal flora; Japanese artichoke.
XX
XX Stachys sieboldii.
XX
XX EP849359-A2.
XX
XX 24-JUN-1998.
XX
XX 18-DEC-1997; 97EP-0122417.
XX
XX 18-DEC-1996; 96JP-0338673.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Oeda K, Mantanabe E;
XX
XX WPI; 1998-324670/29.
XX
XX N-PSDB; AAV40802.
XX
XX New nucleic acid molecule encoding plant raffinose synthetase -
XX capable of producing raffinose, used as food additives with
XX beneficial effects on gastrointestinal flora
XX
XX Claim 1; Page 36-38; 44pp; English.
XX
XX This sequence is the Japanese artichoke raffinose synthetase of the
XX invention. The raffinose synthetase is capable of producing raffinose by
XX combining a D-galactosyl group through an alpha (1-6) bond with a
XX hydroxyl group attached to the carbon atom at position 6 of a D-glucose
XX residue in a sucrose molecule. The DNA can be used to modify metabolism
XX of a host organism by introducing into the host organism or its cell so
XX that the content of the raffinose family oligosaccharides in the host
XX organism or cell is changed. Raffinose family oligosaccharides are useful as
XX food additives with beneficial effects on the gastrointestinal flora.
XX
XX Sequence 587 AA:
SQ
Query Match 59.9%; Score 2529; DB 19; Length 587;
Best Local Similarity 77.0%; Pred. No. 3.2e-230;
Matches 448; Conservative 62; Mismatches 70; Indels 2; Gaps 2;
QY 109 NGDLESETOIVILEKSDSGRPVYLLPVEGPFRTSIOPGDDDFVDVCVSGSSKVVDA 168
DB 2 NGSDLERETQIVLDDKSDS-RPYVILLPLIEGQFRASLOPGVDDFDIEVESGSTVNES 60
QY 169 SFRSMYLXLAGDPPFLVKEAMKIVRHLGTFRLLEKTPRGIVDKFCGCTWDAFYLT 228
DB 61 SFRASLYMHAGDPPFLVDAVVAHNRHGTFRLLLEKTPRGIVDKFCGCTWDAFYLT 120
QY 229 POGVINGVHILVDGCGPPGLVILDDGQSIQHDSPITREKGNQTVAGCQMPRLKFOE 288
DB 121 PHGVMEGVGLVDGCGPPGLVILDDGQSIQHDNDALITREKGMRTSAGCQMPRLKFOE 180

```


PR	30-APR-1998;	98JP-0120551.
PR	04-DEC-1998;	98JP-0345590.
PR	10-DEC-1998;	98JP-0351246.
XX		
PA	(SUMO) SUMITOMO CHEM CO LTD.	

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:02:11 ; Search time 28 seconds
(without alignments)

1161.337 Million cell updates/sec

Title: US-09-675-208-5

Perfect score: 4225

Sequence: 1 MAPSFKNKNGSNVVSFDGLND.....QVWPIDSSSGISVIEYLF 784

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.5	2.7	902	YC47_SCHPO	O14053 schizosacch
2	107	2.5	1174	CIEB_BACTA	O03745 bacillus th
3	106	2.5	1723	AIM1_HUMAN	O094k1 homo sapien
4	105	2.5	730	GLGB_HAEIN	P45177 haemophilus
5	102.5	2.4	564	HEMA_IARUD	P19700 influenza a
6	101.5	2.4	478	FRBK_SALCH	O00330 salmonella
7	101	2.4	514	VG64_HSV11	O00156 ictaluriid h
8	100.5	2.4	712	IRAL_HUMAN	P51617 homo sapien
9	100	2.4	1550	GLTB_SYNY3	P55037 synochocyst
10	99	2.3	770	GLGB_SYNY3	P52981 synochocyst
11	99	2.3	1137	CYG1_CAEEL	O09435 caenorhabd1
12	98	2.3	306	YCBJ_BACSU	P42242 bacillus su
13	97.5	2.3	1095	PIPA_DROME	P13217 drosophila
14	97	2.3	520	HMC5_CRIGR	P13704 citreulius
15	97	2.3	520	HMC5_RAT	P17425 rattus norv
16	97	2.3	859	ENV_ETAV2	P22428 equine infe
17	97	2.3	1002	TAGA_VIBCH	P24019 vibrio chol
18	97	2.3	3061	POLG_PVYHU	O02963 p genome po
19	96.5	2.3	444	YIHL_YEAST	P40512 saccharomyc
20	96.5	2.3	631	DMH1_PARDE	P12293 paracoccus
21	96	2.3	460	ALN_YEAST	P33293 saccharomyc
22	96	2.3	520	HMC5_HUMAN	O01581 homo sapien
23	95.5	2.3	564	HEMA_IAGRE	P16680 influenza a
24	95.5	2.3	346	RUVB_MCMV	O09398 maiz chlor
25	95	2.2	705	CATE_RHIME	O95576 rhizodium m
26	95	2.2	791	LON_CAMJE	O63300 campylobact
27	94.5	2.2	733	AGAI_PEDPE	P43467 pedicoccus
28	94.5	2.2	971	NAC1_RAT	O01728 rattus norv
29	94.5	2.2	1072	CARB_THETN	O818b0 thermoeaer
30	94	2.2	521	IMAL_XENLA	P52170 xenopus lae
31	94	2.2	563	CHIA_SERMA	P07254 serratala ma
32	94	2.2	843	PULA_THEMA	O33840 thermotoga

34	93	2.2	695	1	EEG1_SYNY3	P28371 synochocyst
35	93	2.2	859	1	ENV_ETAV1	P22427 equine infe
36	93	2.2	1047	1	ANPB_BOVIN	P46197 bos taurus
37	93	2.2	3988	1	POLG_BVDVN	P19711 bovine vira
38	92.5	2.2	400	1	EPFU_HERNU	P42477 herpesosiph
39	92.5	2.2	564	1	HEMA_IASE2	P19701 influenza a
40	92.5	2.2	646	1	DHSA_CAEEL	O09508 caenorhabd1
41	92.5	2.2	1356	1	HER1_PODAN	O00808 podospora a
42	92.5	2.2	1431	1	DAPK_HUMAN	P53355 homo sapien
43	92	2.2	859	1	ENV_ETAVC	P32541 equine infe
44	92	2.2	966	1	M172_HUMAN	O14596 homo sapien
45	92	2.2	1065	1	CARB_CLOAB	O971c3 clostridium

ALIGNMENTS

RESULT 1
YC47_SCHPO STANDARD: PRT: 902 AA.
AC 014053:
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein Cl672.07 in chromosome III.
GN SPC1672.07.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetes; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson C.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,
RA Welljens I., Vanhulst E., Rieger M., Schaefer M., Muellet-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehter E., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Mambrt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleure V., Moutier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt G., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez J., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe".
RL Nature 415:871-880(2002).
- - - SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).
- - - This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch)
- - -
EMBL: AL031324; CNA20445.1; - -
InterPro: IPR001680; WD40.


```

Db 650 FCLEKRELSKVAKKRLSDERNLQDPNFRGINRQDRGRSTDTIIGGDDVFKEN 709
Qy 642 PISIEGV--KTFALYLYQAKLILSK-----PSODDLIALDPF--EFELIT 683
Db 710 YTTLEGTPEECPTIYLYQ--KIDESKLAVNRYQLRGYEDSODLEIYIRINAKHEIYN 767
Qy 684 VSPVTKLIQTSLHFAPIGLVNMNTSGAISOVDYDDLS-----SVEIG 727
Db 768 VGTGSLWPLSYE--SPIGCGEPNR--CVPHELMNDLDCSCRDGEKCAHSHHSLLD 824
Qy 728 VKGCGEMRVFASKRPACRIDGEDVYG 753
Db 825 V-GCTDLDQ-----EDLG 835

RESULT 3
ATM1_HUMAN STANDARD: PRT: 1723 AA.
AC 09YAKI; 000296;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Absent in melanoma 1 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97250519; PubMed=9096375;
RA Ray M.E., Miskow G., Su Y.A., Meltzer P.S., Trent J.M.;
RT "ATM1, a novel non-Lens member of the betagamma-crystallin
RT superfamily, is associated with the control of tumorigenicity in human
RT malignant melanoma."
RL Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234(1997).
CC -1- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY
CC EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.
CC -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS
CC 6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RICHIN B-TYPE LECTIN DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: U83116; AAB53792.1; -
DR EMBL: U83115; AAB53791.1; -
DR HSP: P02526; IGAM.
DR Genew: HGNC:356; AIM1.
DR MIM: 601797; -
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR000772; RICHIN_B_lectin.
DR Pfam: PF00030; crystal1; 6.
DR Pfam: PF00652; RICHIN_B_lectin; 3.
DR PRINTS: PRO1367; BGCRCYSTALLIN.
DR SMART: SM00458; RICHIN; 1.
DR SMART: SM00247; XTALBP; 6.
DR PROSITE: PS00325; CRYSTALLIN_BETAGAMMA; 7.
DR PROSITE: PS50331; RICHIN_B_LECTIN; 1.
DR Repeat: LECTIN.
FT DOMAIN 1022 1119 CRYSTALLIN BETA/GAMMA-LIKE 1.
FT DOMAIN 1123 1206 CRYSTALLIN BETA/GAMMA-LIKE 2.
FT DOMAIN 1207 1318 CRYSTALLIN BETA/GAMMA-LIKE 3.
FT DOMAIN 1319 1404 CRYSTALLIN BETA/GAMMA-LIKE 4.
FT DOMAIN 1405 1487 CRYSTALLIN BETA/GAMMA-LIKE 5.
FT DOMAIN 1502 1584 CRYSTALLIN BETA/GAMMA-LIKE 6.
FT DOMAIN 1586 1719 RICHIN B-TYPE LECTIN.

```

```

SQ SEQUENCE 1723 AA; 188646 MW; 7E50F681A627FB09 CRC64;
Query Match 2.5%; Score 106; DB 1; Length 1723;
Best Local Similarity 17.8%; Pred. No. 5.3;
Matches 156; Conservative 127; Mismatches 301; Indels 292; Gaps 41;
Qy 13 VSPDGLNDSSPEFALDGSDFIVNGHSFLSDVPENIVASPSPTSIDKSPVSGCFYGF-D 71
Db 996 ISFGSGS--LSDMILRGS-----YQNKLNPRGKVIYISEPVSSEKCLEVFSD 1042
Qy 72 ASEPDSSRHVVSIGKLDKIDREMSIFREKVMWTHWVGRNGDLE-----SEFOIYLEKS 125
Db 1043 IODCSSWSLSPIVILKVGRCWILYEQPNFECHSIPLIEGELELSGLMGIEDILERHEBA 1102
Qy 126 DSGRPVFLPLIVEGFRSTIQPGDDDPVDCVESSSKVVDASPRSMYLHAAGDP---- 182
Db 1103 ESDK-----PVAIGSIRIVODYRVSHDLFTEPBGLILSSYF-----DDEEM 1147
Qy 183 --FALVKEAMKIVRTHLGFRLIEKTPPGI-----VDKFGWCWDAPYLTVHP-- 229
Db 1148 QGFGVMQKTCG--MKVIMGTWILYEBPGFVGFLEPGEYPLDSFWDFAAYIGSMRPLK 1206
Qy 230 -----OGVIEGVHVLVDGCGCPGLVLIDGW-----OSIGHSDPIRK 267
Db 1207 MGRKVEFPPTDPKVVVYERKPFEGKC---VELETGMCFSVMEGGETEATGDDHLPFS 1262
Qy 268 EGMNQIVAGEQMPCKRLKQENYKFRDYVNPKATGRAGQCKMAFIDLKKEKTEVHEV 327
Db 1263 VGSMAKVLKGIWA-----YEKP-----GFTGHQYLEE--GEYRD---- 1295
Qy 328 YVHNAICGYWGGLRPQGLPREARVIOPLVSPGLQWTMEDLAVDKIVLKVGLVPEKAE 387
Db 1296 --WKA-----WGVTNGEQLG-----RPLTG-----DFSNAHIMVSEKNFGSGSS 1335
Qy 388 EMEGELAHLEKVGIDGVKIDVILHLEMLCEDY-----GGRVLDKAVYKAKMTSINKH 441
Db 1336 IYVLGIVANLKEGTGY--GVYQTSINVLGVAVAYENPDFGEQYIIDKGFYTFEDMGKN 1394
Qy 442 FPGNGVIAEMHCNDPMFLGTEAISLGRGDDFWCIPDPSGDPNGRTWLOGCHMVHNCANS 501
Db 1395 YK-----ISSVQPI-----CLDSTGPPRR-----NOIHLESEP 1423
Qy 502 LMMGNFIHPDMQFOSTHCA--AFHAASRAISGCIYVSDS-----VGKHFDDL 550
Db 1424 QVQGHG-----QSFEETTSQIDDSFTKSCRVSGGSVVYVDGENTFGNLYVEEGHYPL 1478
Qy 551 KKLVLDPDGSILRSEYALPTRDCLFEDPL-----HNGETMLKIMLNKFT 595
Db 1479 SAMGCPPGATFKS---LRFIDVEFSEPTIILFEREDFKKIELNAETV---NLRLS- 1529
Qy 596 GVYIGAFNCG-----GGGW-----CRETRRNCQFSQ 620
Db 1530 ----GFNTQIRSVQVIGIWTYVEGSGYRGROFLSPAEVPMWYEESSGROGSLRPFVQ 1585
Qy 621 YSKRVTSKTNPRDIEWHSSENPISIEGVKTF-----ALYLYQAKLILSKPSQ 668
Db 1586 --KRIYFRLRNKATGLFMFTNG--NLEDKLLRIQWEDVAGADDQIMITY--EGCICKRIAE 1641
Qy 669 DLDIALDPFEFELIYVSPVTKLIQTSLHFAPIGLVNMNTSGAISOVDYDDLSVEIGV 728
Db 1642 DCLLT-----YVSLVTSKSLGLALDQNASQFWSL 1673
Qy 729 KKGCGEMRVFASKRP--ACRIDGEDVGFKKYDQDMVY 753
Db 1674 KSDG--RIYSKLPNLVLDIKG---GTQYDQNHIL 1704

RESULT 4
GGB_HAEIN
ID GGB_HAEIN STANDARD: PRT: 730 AA.
AC P45177;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

```


SQ SEQUENCE 564 AA: 63271 MW: D4010C2548AE3EFC CRC64;
 Query Match 2.4%; Score 102.5; DB 1; Length 564;
 Best Local Similarity 18.6%; Pred. No. 2;
 Matches 110; Conservative 61; Mismatches 186; Indels 233; Gaps 26;

```

OY 7 NGGSNVVSEFDGLND-----MSSPFAIDGSDPTVNGHSLSDPENIVASPSPTSIDKS 60
DB 80 NGALGSPGDHNLGAMWDFIERPTAVD-----TCYPFVDP-----YQSIRSI 123
OY 61 PVSVCCEFGFADSEPDSSRHVSIK---LKDIRFMSIFEFKAWWTTTHWVR----- 108
DB 124 LANNKRF-EFIAEEFQWMTVKONGKSGACKRANVNDFFR-RLNWLTKSGNAYPLONLTK 181
OY 109 -NGGDL-----ESETQIVLEKSDSGR-----PYVFLPIVEGP-- 141
DB 182 VANGDYARLYTNGVHHPTSDTQTNLYKNNPGRVYVSTKTSQTSVYPNIGSRPLVRGSG 241
OY 142 ---FRTSIOPGD-----DDFVDCVESGSS 163
DB 242 RISFYWTIVEPDGLVFNITGNLIAPRGHYKLNOKKSTILMTAVPISGCVSKCHTNRGS 301
OY 164 KYVDASFNSMLYLHAGDDPFALVKEAMKIVRHLGLTFLLEKTPPGIYDKRGWCTWDAF 223
DB 302 ITTTRPFONISHSIGDCRYKQSLKLTAT---GMRNIPKATRGLE----- 346
OY 224 YLTVPQGVIEGVRLVLDGCPPLGLVLDGMSQISGHSDPTTKKGMNQTVAGEQMPRL 283
DB 347 -----GAIAG-----FIENGNGL----- 360
OY 284 LKFOENYKRDYVNPKATGPRAGQKMAFIDELKGEFTVEHYVMAHLCGYWGLRQ 343
DB 361 --IDGWYGR-HQNEGTGTADLKSTQAIQDINGKLNRL----- 398
OY 344 VGLPEARIQYVLSGLQMTMEDLAVDKIVLHKVGLVPEKAEEMVEGLHAHLEKVID 403
DB 399 -----IEKTNEK-----YHQI-----EKEFEQVEIRIDLEKY-VE 428
OY 404 GVKIDV-----IHLEMLCEDYGRVDAKAYKAMTKSI-----NKHFKNGVISMHE 453
DB 429 DTKIDLMSTNAELVALLEMOH--TIDVTSEMKKLEFERKRLRENAEDKGNCFEIFRD 486
OY 454 CNDFMFLTEAISLGRVDDFMCTDPSGDPNGTFWLGCHMYHACAND-SLWM 504
DB 487 CONNCT--ESTRNGTYDHDYRDEAI---NNFQIQGVKLTGYKDILMI 532
  
```

RESULT 6
 REBK_SALCH STANDARD: PRT: 478 AA.
 ID 000330;
 DT 01-OCT-1994 (rel. 30, Last sequence update)
 DT 01-OCT-1994 (rel. 30, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Phosphomannomutase (EC 5.4.2.8) (PMM).
 GN MANB OR REBK.
 OS Salmonella cholerae-suis (Salmonella enterica).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=591;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=M67;
 RX MEDLINE=92349966; PubMed=1379320;
 RA Brown P.K., Romana L.K., Reeves P.R.:
 RT "Molecular analysis of the rfb gene cluster of Salmonella serovar
 RT muenchen (strain M67): the genetic basis of the polymorphism between
 RT groups C2 and B.";
 RL Mol. Microbiol. 6:1385-1394(1992).
 CC -I- FUNCTION: INVOLVED IN GDP-MANNOSE BIOSYNTHESIS WHICH SERVES AS
 CC THE ACTIVATED SUGAR NUCLEOTIDE PRECURSOR FOR MANNOSE RESIDUES
 CC IN CELL SURFACE POLYSACCHARIDES. THIS ENZYME PARTICIPATES IN
 CC SYNTHESIS OF THE LPS GROUP C2 O ANTIGEN.

CC -I- CATALYTIC ACTIVITY: D-mannose 1-phosphate -> D-mannose 6-phosphate.
 CC -I- PATHWAY: GDP-MANNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
 CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -I- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

```

DR EMBL, X61917; CAA43916.1; -
DR PIR, S22622; S22622.
DR InterPro: IPR001485; PG/PMM_mutase.
DR Pfam: PF00408; PGM_PMM; 1.
DR Pfam: PF02878; PGM_PMM_I; 1.
DR Pfam: PF02879; PGM_PMM_II; 1.
DR Pfam: PF02880; PGM_PMM_III; 1.
DR PROSITE: PS00710; PGM_PMM; 1.
KW Lipopolysaccharide biosynthesis; Isomerase; Transmembrane;
KW Phosphorylation.
FT TRANSMEM 30 46
FT TRANSMEH 265 284
FT ACT_SITE 111 111
FT ACT_SITE 111 111
  
```

SQ SEQUENCE 478 AA: 52813 MW: 982E5B885083893 CRC64;
 Query Match 2.4%; Score 101.5; DB 1; Length 478;
 Best Local Similarity 21.8%; Pred. No. 1.9;
 Matches 59; Conservative 30; Mismatches 79; Indels 103; Gaps 14;

```

OY 349 EARVIOPVLSPLQMTMEDLAVDKIVLHKVGLVPEKAEEMVEGLHAHLEKVIDKID 408
DB 141 DMSFQPKLE---QLTISTIAARNYILRTYSLFPMPLKRRKRGITDH----- 185
OY 409 VHLLEMLCEDYGRVDAKAYKAMTKSINKHFGKNGVJASNEHCNDFMFLTEAISLG 468
DB 186 -----SSAGR-DLVKTLFLKML-----GATVVSILARSDEFPIDTEAVS-- 222
OY 469 RVGDD-----FVCTDPSG-----DPNGTFWLOG-----CHM----- 494
DB 223 --EDDRNKAITWAKKYQLDALIFSTDGDRPLADEYGN-WLREGDILGLCSLELADAV 279
OY 495 ---VHCANDSLWKGNT-HPDMDMFQSTHPCAFHAASRAISGPIYSDSVGKHNPDLL 550
DB 280 AIPVSC-NSTISSGNFKNVETKISPIYIAFAKLSA-----NYNCT 322
  
```

RESULT 7
 VG64_HSVII STANDARD: PRT: 514 AA.
 ID VG64_HSVII
 AC 000156;
 DT 01-DEC-1992 (rel. 24, Created)
 DT 01-DEC-1992 (rel. 24, Last sequence update)
 DT 01-DEC-1992 (rel. 24, Last annotation update)
 DE Hypothetical gene 64 protein.
 GN 64.
 OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Ictalurid herpes-like viruses.
 OX NCBI_TaxID=10401;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Auburn 1;
 RX MEDLINE=92087490; PubMed=1727613;
 RA Davison A.O.:
 RT "Channel catfish virus: a new type of herpesvirus.";


```

AC P55037.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferredoxin-dependent glutamate synthase 1 (EC 1.4.7.1) (Fd-GOGAT).
GN GLTB OR SLI1502.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95244836; PubMed=7727752;
RA Navarro F., Chavez S., Candau P., Florencio F.J.;
RT "Existence of two ferredoxin-glutamate synthases in the
RT cyanobacterium Synchocystis sp. PCC 6803. Isolation and insertional
RT inactivation of gltb and glts genes.";
RL Plant Mol. Biol. 27:753-767(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kikura T.,
RA Hoshouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-116(1996).
CC -1- CATALYTIC ACTIVITY: 2 L-glutamate + 2 oxidized ferredoxin = L-
CC glutamine + 2-oxoglutarate + 2 reduced ferredoxin.
CC -1- COFACTOR: BINDS A 3FE-4S CLUSTER; FAD AND FMN.
CC -1- PATHWAY: GLUTAMINE SYNTHETASE/GOGAT PATHWAY WHICH IS INVOLVED
CC IN THE ASSIMILATION OF AMMONIA.
CC -1- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-21 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X80485; CAA56552.1;
CC EMBL: D90902; BAA17018.1;
CC InterPro: IPR002489; DUF14.
CC InterPro: IPR002932; Glu_synthase.
CC Pfam: PF01493; DUF14.1.
CC Oxidoreductase: Iron-sulfur; 3Fe-4S; Flavoprotein; FAD; FMN;
CC Glutamate biosynthesis: Complete proteome.
CC DOMAIN
CC 43 393 GLUTAMINE AMIDOTRANSFERASE (POTENTIAL).
CC NP_BIND 1057 1154 FMN (BY SIMILARITY).
CC METAL 1150 1150 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
CC METAL 1156 1156 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
CC METAL 1161 1161 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
CC SEQUENCE 1550 AA; 169071 MW; 07AA64DB32296943 CRC64;
Query Match 2.4%; Score 100; DB 1; Length 1550;
Best Local Similarity 19.9%; Pred. No. 14;
Matches 136; Conservative 89; Mismatches 224; Indels 234; Gaps 38;

```

```

Db 498 -FEEELRILLAMG-RDGEAIGSMGADTPLAVLSDBKRLLYNYEQLEAQTNPIDSI 555
QY 266 TKE---GMNQVAGE-----QMPCTL-----LKEQENYKRDVYNPKA 300
Db 556 REEITTSATTITGGGCGNLDPRESCRLIELKTPILTNEDLAKLALDDDEKSVTLDL 615
QY 301 TGPRAGQGMKAFIDELGGEK-----YVEHYVMHALCGYWG----- 338
Db 616 FDPNGEGAGLKTALDNLFTADQALISOGANLITLSDQVSAEKAL-PALLAVSGLHHLL 674
QY 339 ---GLRQVPGRLPEARVLOPVLSBGLQMTMEDLAVDKYLAVGLVPEPKAEEMEGHLA 395
Db 675 IANGRTKV-GL-----VLESGEPREVNHFV--LLGYCGAINPYLAETLIDGMIA 723
QY 396 HLEKVGIGDKVIDVHLEMLCEDYGGRVDLKAVYKATKTSINHFNGNGVIASMECN 455
Db 724 -----EGLVNVDH--KTAC-----KNYKATKGVITVASKIGISTQSTRG 764
QY 456 DEMPLGTEAISGR-VGDDFWCTDPSGDPNGTFWLOGCHMVHCANDSLMGNFIHPDWM 514
Db 765 AQIF--EAVGLNGSVIDEVPCRTSSR-----IQSDLGVIAGEALIR----- 804
QY 515 FQSTHPCAFHAAISRAISGPIVYSDSVKHF-----DLKRLVLPDGTI 560
Db 805 -----HONAFAPRPGDLHTLDVGEYQWRKDGEHLFSPOTIHLLQRAVEGNYE 854
QY 561 LRSEYVAL-----PTROCL-FED-----PLHNGETMLKIMNLNKF-TGVICAFNCG 605
Db 855 LKQYALVNEONOKFFTLRGILDFQDRESIPLEVEPEIAL--MKRFYGMSTV----- 907
QY 606 GWCETRRNOCFSQSKVTSKTN-----PKDIEHSGENPISIEGVKFA----- 952
Db 908 GSISAEHESLAIAM--NIGGKSTMGEGEDPERFTWINDGDSKNSAIKQVASGRGV 965
QY 653 --LYIQAKKIL-----SKPSQ 668
Db 966 TSLYLSQAKKEIQIKMAOKAPGE 988
RESULT 10
GLGB_SYNY3
ID GLGB_SYNY3 STANDARD; PRT; 770 AA.
AC P52981;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (glycogen branching
DE enzyme).
GN GLGB OR SLI0158.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiyama M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
CC glycogen.
CC -1- PATHWAY: Glycogen biosynthesis: third step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```

OY 298 PRATGPAQOKKAKAFIDELKGEFKEVHEVYVWALCGYWGGLRPOVPLPEARVQPV 357
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 814 PRANICISOMKGLVS-----KOTNLMDHVF--NMDEXTSTLEEET-----EERT----- 857
OY 358 SEPOLMTMEDLAVDKVLKVLGVLPPEKABEMTEGHAHLEKVGIDGVKI---DVTH--L 412
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 858 ---KELTLEKKKKD-ILLSR--MLPKQVABRLKAG--QVPEPGFSVYVFFSDVFKFTI 909
OY 413 LEMLCEDYCGRYDLAKAVYKAMTKSINKHFKNGVGLASHECHDFELGTAEATSLGRVD 472
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 910 LMSKCSPF-QTVNLMDLYSNEDTITEQH---GV-----YKVESIGD 947
OY 473 DEWCTPSPDSDNGTFWLGQCHVYHCANDSLMGNFTHPDMDFOSTHPCAAFRASR--- 529
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 948 GYLCVSGPLTRNGY-----AHIKQIVDMSLKEMEY-----CKSFNTHPLPRE 989
OY 530 -----AISGGPIYVSDVSGKHNFDLLKLVLPDGSILNSEYALPTROCLFEDPLHNGE 583
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 990 MWELRIGVNSGPC-VAGVVG-----LSMP-----RYCLEFQDTVNTAS 1025
OY 584 TM 585
DB 1026 RM 1027

RESULT 12
YCBJ_BACSU
ID YCBJ_BACSU STANDARD: PRT: 306 AA.
AC P42842:
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ycbj.
GN YCBJ
OS Bacillus subtilis.
OC Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168:
RX MEDLINE=98044033; PubMed=9384377;
RA Kunest F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Besterio M.G., Bessieres P., Bolotin A., Borcher S.,
RA Bortiss R., Bourisier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillett S., Bruschel C.V., Caldwell B., Capriano V., Carter N.M.,
RA Chol S.K., Codant J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlan K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Glaeser P., Goffeau A., Goldthly E.J., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Goldthly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holseppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasanara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Kohnigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Presenc A., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiuchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Taccoti E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaret A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Wetzeneeger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
```

```

RN [2]
RP SEQUENCE OF 98-306 FROM N.A.
RC STRAIN=168:
RX MEDLINE=95219079; PubMed=7704254;
RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;
RT "Determination of a 21548 bp nucleotide sequence around the 24
RT degrees region of the Bacillus subtilis chromosome."
RL Microbiology 141:269-275(1995).
CC -1 SIMILARITY: TO AMINOGLYCOSIDE PHOSPHOTRANSFERASES AND SPECIFICALLY.
CC TO VIOMYCIN PHOSPHOTRANSFERASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC
CC EMBL: 299105; CAB12046.1;
CC DR EMBL: D30808; BAA06474.1;
CC DR Subtilist: BG1165; ycbj.
CC KW Hypothetical protein, Transferase; Kinase; ATP-binding;
CC Complete proteome.
CC FT ACT_SITE 204
CC FT SITE 204
CC SQ SEQUENCE 306 AA; 34481 MW; 5EF0FD9DEC085977D CRC64;

Query Match 2.3%; Score 98; DB 1; Length 306;
Best local Similarity 20.1%; Pred. No. 1.9;
Matches 54; Conservative 33; Mismatches 84; Indels 98; Gaps 11;

OY 120 VILEKSDSGRPVY-----FL-----LPIV 138
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 40 VYAKNDTGMPWLKRRPRSDYVERASAGITLAFILANULTADVPRMRTHTPELVPL 99
OY 139 EGFPRST-----QPDDEY---DVCVESGSSKVVDSFRSMKLYLAGD 181
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 100 KGPRAAGIDLEOKQYVWNNMDHPSPDFFVTLADIAE-----LH-GTD 142
OY 182 PFALVEAMKIVYTHIGTFRLLEKTPPGIVDFGCT--WDAFYLTVPQVIEGVRL 239
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 143 QISAGSGIEVIRPE--DRQMTADSMVDVKNKLGSTLMEWQKVVDDAVWPQFSSL 200
OY 240 VQGCPCPGVLIDDCWQSG-----HSDPTKCGMGTVAAGQPCRLKFOENYKF 292
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 201 IRHDLRPHILIDQNRVYGLDWTAKVADPRKDVLVQTINGEKETARLLLEYDQ----- 257
OY 293 RDVYNPKATGPRAQOKKAFIDELKGEF 321
      | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: | 11: |
DB 258 -----AGGRIMAK-MOEHISEMQAAY 277

RESULT 13
PIPA_DROME
ID PIPA_DROME STANDARD: PRT: 1095 AA.
AC P13217:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 1-phosphatidylinositol-4,5-bisphosphate phospholipase
DE (EC 3.1.4.11) (phospholipase C-specific phospholipase C).
GN NORPA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88311074; PubMed=2457447;
RA Bloomquist B.T., Shortridge R.D., Schneuwly S., Perdew M.H.,
RA Monteil C., Steller H., Rubin G., Pak W.L.;
RT "Isolation of a putative phospholipase C gene of Drosophila, norPA,
```

```

RT and its role in phototransduction."
RL Cell 54:723-733(1988).
CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES
CC DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS
CC MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
CC C ENZYMS. NORA IS INVOLVED IN PHOTOTRANSDUCTION.
CC -1- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-
CC bisphosphate + H(2)O = D-myo-inositol 1,4,5-trisphosphate +
CC diacylglycerol.
CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS OF
CC PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03138: AAA28724.1; -
CC PIR: A31225: A31225.
CC HSSP: P10688: 10A5.
CC Flybase: FBgn0004625; noirpa.
CC InterPro: IPR000008; C2.
CC InterPro: IPR001192; PI_PLC.
CC InterPro: IPR000909; PI_PLC_xdom.
CC InterPro: IPR001711; PI_PLC_X.
CC Pfam: PF00168; C2: 1.
CC Pfam: PF00387; C1-PLC-X; 1.
CC Pfam: PF00388; PI-PLC-X; 1.
CC PRINTS: PR00390; PHPLIPASEC.
CC ProDom: PD001202; PI_PLC_Y; 1.
CC SMART: SM00239; C2: 1.
CC SMART: SM00148; PLCX; 1.
CC SMART: SM00149; PLCX; 1.
CC PROSITE: PS50004; C2_DOMAIN.2; 1.
CC PROSITE: PS50007; PIPLC_X_DOMAIN; 1.
CC PROSITE: PS50008; PIPLC_Y_DOMAIN; 1.
CC HydroLase: Lipid degradation; Vision; Transducer.
CC FT DOMAIN 320 469 DOMAIN X.
CC FT 550 666
CC FT DOMAIN 673 771 C2 DOMAIN.
CC FT ACT_SITE 334 381 BY SIMILARITY.
CC FT ACT_SITE 381 381 BY SIMILARITY.
CC SEQUENCE 1095 AA: 124822 MW: 2D945BF0ACB69B3 CRC64:
Query Match 2.38; Score 97.5; DB 1; Length 1095;
Best Local Similarity 19.48; Pred. No. 13;
Matches 109; Conservative 69; Mismatches 210; Indels 175; Gaps 26;

```

```

Db 500 AGKPEEAAAAPAPAEAAAEEGCGGAEAAEAANYSGSTTVNHPMLSMV-NYAQ 558
QY 423 RVDLAKAYVYMTSIRKHKRGKGVISMHC--NDPMFGTEALISGRVGDDEWCTDPS 480
Db 559 PIKF-QGFDAIEKNINHNNSFESAGMYLKOSIDFVYMKRQMSRIYPKCTRADSS 617
QY 481 GDPNGTEWLOGCHVHCANDS-----LMMGNFTH-----PDMDWFGS 517
Db 618 NYMPQVFMNQGCVWVNLFPQSDPLPMQLNGKFFENGGCVLLKPDPMRADKDFDFAD 677
QY 518 -----THPCAAHAAKRAISGPT--YV-----SSVCK-----H 545
Db 678 APVDGVIAACSVYVIAQGLFSDKKVGYEVDVFMGLPSDTYKKEFRTLVAANGLNPLY 737
QY 546 NFD--LLKLVLPDGSILRSEY 566
Db 738 NEDPFRKVVLPDLAVLRGVY 760

RESULT 14
HMCS_CRIGR
ID HMCS_CRIGR STANDARD: PRT: 520 AA.
AC P13704;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 4.1.3.5) (HMG-CoA
DE synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN HMGS1 OR HMGS.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=66140166; PubMed=2869035;
RA Gill G., Goldstein J.L., Slaughter C.A., Brown M.S.;
RT "Cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase from the
RT hamster. I. Isolation and sequencing of a full-length cDNA."
RL J. Biol. Chem. 261:3710-3716(1986).
CC -1- FUNCTION: THIS ENZYME CONDENSES ACETYL-CoA WITH ACETOACETYL-CoA
CC TO FORM HMG-CoA, WHICH IS THE SUBSTRATE FOR HMG-CoA REDUCTASE.
CC -1- CATALYTIC ACTIVITY: (S)-3-hydroxy-3-methylglutaryl-CoA + CoA =
CC acetyl-CoA + H(2)O + acetoacetyl-CoA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-CoA PRIOR TO THE
CC SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOLIDS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE HMG-CoA SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L00334; AAA37076.1; -
CC EMBL: L00326; AAA37076.1; JOINED.
CC EMBL: L00327; AAA37076.1; JOINED.
CC EMBL: L00328; AAA37076.1; JOINED.
CC EMBL: L00329; AAA37076.1; JOINED.
CC EMBL: L00330; AAA37076.1; JOINED.
CC EMBL: L00331; AAA37076.1; JOINED.
CC EMBL: L00332; AAA37076.1; JOINED.
CC EMBL: L00333; AAA37076.1; JOINED.
CC PIR: A25332; A25332.
CC InterPro: IPR000590; HMG-CoA_synth.
CC Pfam: PF01154; HMG-CoA_synth; 1.
CC PROSITE: PS01226; HMG-CoA_SYNTHASE; 1.

```

KW Lyase; Cholesterol biosynthesis; Multigene family.
 FT ACT_SITE 129 129 POTENTIAL.
 SO SEQUENCE 520 AA; 57318 MW; 2DACAEE7DEE5D6BB CRC64;

Query Match 2.3%; Score 97; DB 1; Length 520;
 Best Local Similarity 18.8%; Pred. No. 5;

Matches 80; Conservative 63; Mismatches 115; Indels 168; Gaps 21;

```

OY 378 VGLVPE-----KAE-EMEGLAHLEKVGIDGVKID-----VIHL 412
DB 17 VGIVALEIFPPSOYVDQALEKTDGVDAKRYITIGLQAKMGCTDIEDINSICLITYVQNL 76
OY 413 LE--MLCEDYGGKRVDLAKAYKAMTKSINKHF-----KNGVYASMEHCNDFMLGTEPA 464
DB 77 MERNLSLSDYICIGLEVTGETITIDSKSVKSNLMQLFEESGNTDIEGIDPTN--ACYGCTAA 135
OY 465 ISLGRVGDDPFWCTDPSGDNGTFMLQGCCHVHCANDSLMKGNF---IHPDMDMFQ--STH 519
DB 136 VF-----NAVMWIE-----SSWDGRVALVAGDIAIYATGNAR 169
OY 520 PCAAFHAAASRAISGPIYVSD-----SVGKHNFDLKLLVLPDGSILRSEYVALPTROCL 574
DB 170 PGCGVAAVALLLIGPNAPLFPDRGLRGTTHQHAIDFK-----PD--MLSEY----- 213
OY 575 FEDPLHNGETMLKIMNLNKFYIGAFNCOGGGWCRTRRNOCFSQYSKRVTSKTNPKDI 634
DB 214 ---PIVDGKLSIQ-----CYLSALDRCSYVRKKIRA----- 242
OY 635 EHHSGENPISEGVKTFALYLYOAKKLLSKPSQDIDIALDPFEFLITVSPYTKLIOTS 694
DB 243 QW-----QKEGNDNDFTLNDFGF--MISHSPYCKLVQKS 274
OY 695 LHFAPIGLVNMLNTSGAISOVDYDDLSVEIGKCGGEMRVAFSKPRACRIDGEDVGF 754
DB 275 L--ARMLNDFLN-----DONRDKNSIYSGLEAFGDVKL-----EDTYF 311
OY 755 KYDQDQ 760
DB 312 DRDVER 317

RESULT 15
HMGCS_RAT STANDARD: PRT: 520 AA.
AC P17425;
DT 01-AUG-1990 (rel. 15, Created)
DT 01-AUG-1990 (rel. 15, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 4.1.3.5) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN HMGC1 OR HMGC5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX MEDLINE=90301491; PubMed=1972979;
RA Ayte J., Gil-Gomez G., Hegardt F.G.;
RT "Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-hydroxy-3-methylglutaryl coenzyme A synthase."
RL Nucleic Acids Res. 18:3642-3642(1990).
CC -1- FUNCTION: THIS ENZYME CONDENSES ACETYL-CoA WITH ACETOACETYL-CoA TO FORM HMG-CoA, WHICH IS THE SUBSTRATE FOR HMG-CoA REDUCTASE.
CC -1- CATALYTIC ACTIVITY: (S)-3-hydroxy-3-methylglutaryl-CoA + CoA -> acetyl-CoA + H(2)O + acetoacetyl-CoA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-CoA PRIOR TO THE SYNTHESIS OF STEROIDS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE HMG-CoA SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X52625; CAA36852.1; .
 DR PIR: S12736; S12736.
 DR InterPro: IPR000590; HMG-CoA-synt.
 DR Pfam: PF01154; HMG-CoA-synt. 1.
 DR PROSITE: PS01226; HMG-CoA-synthase: 1.
 KW Lyase; Cholesterol biosynthesis; Multigene family.
 FT ACT_SITE 129 129 POTENTIAL.
 SO SEQUENCE 520 AA; 57433 MW; CB213A27B0C177CB CRC64;

Query Match 2.3%; Score 97; DB 1; Length 520;
 Best Local Similarity 18.8%; Pred. No. 5;

Matches 80; Conservative 62; Mismatches 117; Indels 166; Gaps 21;

```

OY 378 VGLVPE-----KAE-EMEGLAHLEKVGIDGVKIDYIHLLEMLCEDYGGKRVDLA 427
DB 17 VGIVALEIFPPSOYVDQALEKTDGVDAKRYITIGLQAKMG-----FCTD--REDYN 66
OY 428 KAYKAMTKSINKHFKNGVYASMEHCNDFMLGTEPAI-----SLGRVGDDPFWCTD 478
DB 67 SLCLTYVQKLMERNLSYDICIQLE-----VGFETIIDSKSVKSNLMQLFEESGNTD 119
OY 479 PSGDP-----NGTFMLQGCCHVHCANDSLMKGNF---IHPDMDMFQ--THP 520
DB 120 IEGIDTTNACYGTAAVFNAVWIE-----SSWDGRVALVAGDIAIYASGNAR 170
OY 521 CAAAFHAAASRAISGPIYVSD-----SVGKHNFDLKLLVLPDGSILRSEYVALPTROCL 575
DB 171 TCGCVAAVALLLIGPNAPVIFDRGLRGTTHQHAIDFK-----PD--MLSEY----- 213
OY 576 EDDLHNGETMLKIMNLNKFYIGAFNCOGGGWCRTRRNOCFSQYSKRVTSKTNPKDIE 635
DB 214 --PVVDGKLSIQ-----CYLSALDRCSYVRKKIRA-----Q 243
OY 636 WHSGENPISEGVKTFALYLYOAKKLLSKPSQDIDIALDPFEFLITVSPYTKLIOTS 695
DB 244 W-----QKEGNDNDFTLNDFGF--MISHSPYCKLVQKSL 275
OY 696 HFAPIGLVNMNTSGAISOVDYDDLSVEIGKCGGEMRVAFSKPRACRIDGEDVGF 755
DB 276 --ARMLNDFLN-----DONRDKNSIYSGLEAFGDVKL-----EDTYFD 312
OY 756 YDQDQ 760
DB 313 RDVER 317

```

Search completed: April 4, 2003, 11:11:49
 Job time : 36 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:09:51 ; Search time 29 seconds
(without alignments)
795.433 Million cell updates/sec

Title: US-09-675-208-5
Perfect score: 4225
Sequence: 1 MAPSFKNNGSNVVSFDGLND.....QVPWPIIDSSGSGISVIEYLF 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUTS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4225	100.0	784	4 US-08-846-234-5	Sequence 5, Appl1
2	181	4.3	30	4 US-08-846-234-1	Sequence 1, Appl1
3	109	2.6	1390	4 US-08-770-544-2	Sequence 2, Appl1
4	107.5	2.5	1584	4 US-09-251-645-6	Sequence 6, Appl1
5	107	2.5	487	4 US-09-004-838-101	Sequence 101, Appl
6	107	2.5	488	4 US-08-040-751-3	Sequence 47, Appl
7	107	2.5	1174	1 US-08-291-368-2	Sequence 3, Appl1
8	107	2.5	1174	1 US-08-291-368-2	Sequence 2, Appl1
9	107	2.5	1174	2 US-08-962-190-2	Sequence 2, Appl1
10	107	2.5	1174	5 PCT-US95-10310-2	Sequence 2, Appl1
11	107	2.5	1174	6 5164180-4	Patent No. 5164180
12	105	2.5	19	4 US-08-846-234-2	Sequence 2, Appl1
13	102	2.4	478	4 US-09-819-993-2	Sequence 2, Appl1
14	100	2.4	518	4 US-09-819-993-5	Sequence 5, Appl1
15	100	2.4	520	1 US-08-305-505-6	Sequence 6, Appl1
16	100	2.4	1276	1 US-08-717-515-8	Sequence 8, Appl1
17	98.5	2.3	770	1 US-08-525-654A-1	Sequence 1, Appl1
18	98.5	2.3	771	1 US-08-525-654A-3	Sequence 3, Appl1
19	98	2.3	485	4 US-09-004-838-131	Sequence 131, Appl
20	97	2.3	472	4 US-09-004-838-48	Sequence 48, Appl
21	97	2.3	472	4 US-09-004-838-103	Sequence 103, Appl
22	97	2.3	520	1 US-08-305-505-4	Sequence 4, Appl1
23	97	2.3	520	1 US-08-305-505-5	Sequence 5, Appl1
24	96	2.3	520	1 US-09-819-993-4	Sequence 4, Appl1
25	96	2.3	772	4 US-09-134-078-28	Sequence 28, Appl
26	95.5	2.3	712	1 US-08-587-889-2	Sequence 2, Appl1
27	95.5	2.3	712	1 US-08-980-060-5	Sequence 5, Appl1

28	95.5	2.3	712	4 US-09-307-185-5	Sequence 5, Appl1
29	95.5	2.3	712	5 PCT-US96-09193-2	Sequence 22, Appl1
30	95	2.2	600	4 US-09-388-743-22	Sequence 2, Appl1
31	94	2.2	750	4 US-09-364-970-8	Sequence 8, Appl1
32	93.5	2.2	540	4 US-08-687-580B-7	Sequence 7, Appl1
33	93.5	2.2	600	4 US-09-423-468A-13	Sequence 13, Appl1
34	92.5	2.2	1423	4 US-08-810-712-10	Sequence 10, Appl1
35	92	2.2	280	2 US-08-403-853-10	Sequence 10, Appl1
36	92	2.2	649	4 US-09-066-047-5	Sequence 5, Appl1
37	92	2.2	820	4 US-09-313-677-21	Sequence 21, Appl1
38	92	2.2	926	4 US-09-313-677-2	Sequence 2, Appl1
39	92	2.2	933	4 US-09-313-677-19	Sequence 19, Appl1
40	92	2.2	967	4 US-09-313-677-17	Sequence 17, Appl1
41	90.5	2.1	504	4 US-09-134-001C-2980	Sequence 1980, Ap
42	90.5	2.1	592	4 US-09-423-468A-15	Sequence 15, Appl
43	89.5	2.1	1854	4 US-09-004-838-108	Sequence 108, Appl
44	89	2.1	559	4 US-09-242-690A-15	Sequence 15, Appl
45	89	2.1	559	4 US-09-298-924-6	Sequence 6, Appl1

ALIGNMENTS

```

RESULT 1
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-846-234-5

Query Match      100.0%; Score 4225; DB 4; Length 784;
Best Local Similarity 100.0%; Pred No. 0;
Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 MAPSFKNNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVPENIVASPSPTSIDKS 60
      |||
Db      1 MAPSFKNNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVPENIVASPSPTSIDKS 60
      |||
QY      61 PVSVCPCVGFDAEPPSRHVSIGKLDKIDRFNSIFRKKVMTTHWVGRCGDLESEFOIV 120
      |||
  
```

```

Db      61 PVSVCNFGFSDASEPDSRRHVSIKLIKIDREMSIFREKYVMWTHWVGRNGDLESTQIV 120
QY      121 ILEKSDSGRPYFLLPIYVGPFRTSIQPGDDDFVDYCVESGSSKVVYDASFRSMLYLHAD 180
Db      121 ILEKSDSGRPYFLLPIYVGPFRTSIQPGDDDFVDYCVESGSSKVVYDASFRSMLYLHAD 180
QY      181 DFFALYKEMKIVRHHGLTFRLLLEKTPPGIYDKFEGMCTWDAFYLTVHQVIEGRHLV 240
Db      181 DFFALYKEMKIVRHHGLTFRLLLEKTPPGIYDKFEGMCTWDAFYLTVHQVIEGRHLV 240
QY      241 DGGCPGGLVLIIDGMSISIHSDSPITKEGNOTVAGEQPCRLKQOEYKFRDYVNPKA 300
Db      241 DGGCPGGLVLIIDGMSISIHSDSPITKEGNOTVAGEQPCRLKQOEYKFRDYVNPKA 300
QY      301 TGPRAQOKMKAFIDELKGEFTVEHVYVWALCGYWGSLRPQVGLPEARVIOPLVSPG 360
Db      301 TGPRAQOKMKAFIDELKGEFTVEHVYVWALCGYWGSLRPQVGLPEARVIOPLVSPG 360
QY      361 LQMTMEDLAVDKIVLHKVGLVPPERKAEMEYEGILHAHLEKVGIDYKIDYIHLEMLCEBY 420
Db      361 LQMTMEDLAVDKIVLHKVGLVPPERKAEMEYEGILHAHLEKVGIDYKIDYIHLEMLCEBY 420
QY      421 GGRVLDKAYYKAMTKRSINKHFKNGVITASMEHCNDPMFLGTETALSIGVGDPMCTDPS 480
Db      421 GGRVLDKAYYKAMTKRSINKHFKNGVITASMEHCNDPMFLGTETALSIGVGDPMCTDPS 480
QY      481 GDPNGTFWLOGCHMYVHCANDSLMNGFIHPDMDMFQSTHPCAFHAAASRAISGGPIYVSD 540
Db      481 GDPNGTFWLOGCHMYVHCANDSLMNGFIHPDMDMFQSTHPCAFHAAASRAISGGPIYVSD 540
QY      541 SVGKHNFDLKLVLPGDGLRSEYVIALPTROCLFEDPLHNGETMKNLNKFTGVITA 600
Db      541 SVGKHNFDLKLVLPGDGLRSEYVIALPTROCLFEDPLHNGETMKNLNKFTGVITA 600
QY      601 FPGCGGCMCRERTRNOCFOYSKRVTSKTPKDIEHSGENPISIEGVTFPALYLYQAAK 660
Db      601 FPGCGGCMCRERTRNOCFOYSKRVTSKTPKDIEHSGENPISIEGVTFPALYLYQAAK 660
QY      661 LILSRPSODLIALDPFEELITVSPVKLIQTSIHFAPIGLVNMMLTSGAIOGVYDGD 720
Db      661 LILSRPSODLIALDPFEELITVSPVKLIQTSIHFAPIGLVNMMLTSGAIOGVYDGD 720
QY      721 LSSVEIGVGCCEMVFASKPRACRIDGEDYGFYDQOMVYVQVWPIIDSSGGISVI 780
Db      721 LSSVEIGVGCCEMVFASKPRACRIDGEDYGFYDQOMVYVQVWPIIDSSGGISVI 780
QY      781 EYLF 784
Db      781 EYLF 784

```

RESULT 2
US-08-846-234-1
Sequence 1, Application US/08846234
Patent No. 6166292

GENERAL INFORMATION:
APPLICANT: OSUMI Chieko
APPLICANT: NOZAKI Jinshi
APPLICANT: KIDA Takao
TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,234
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-413-3000
TELEFAX: (703)-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-846-234-1

Query Match 4.3%; Score 181; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.6e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 FGCWCTDAFYLYVHPQGVIEGRHLVDGCG 244
Db 1 FGCWCTDAFYLYVHPQGVIEGRHLVDGCG 30

RESULT 3
US-08-770-544-2
Sequence 2, Application US/08770544
Patent No. 5907085

GENERAL INFORMATION:
APPLICANT: Gonssalves, Dennis
TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS PROTEINS AND
THEIR USES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,544
FILING DATE:
CLASSIFICATION: A35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60009008
FILING DATE: 21-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/621
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1390 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-770-544-2

Oy 738 ---ASKKPRACRIGEDVGFKYDQDMVVVQVFWPIDSSGGISVLEY 782
 Db 980 SYDAOGRIVKTEQDAQYATFEYDMVGRLLITTTTKDTTSLSQLVTKLEY 1027

RESULT 5

US-09-004-838-101
 ; Sequence 101, Application US/09004838
 ; Patent No. 6350933
 ; GENERAL INFORMATION:
 ; APPLICANT: Michelmore, Richard W.
 ; APPLICANT: Meyers, Blake
 ; TITLE OF INVENTION: Procedures and Materials for
 ; NUMBER OF SEQUENCES: 140
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/004.838
 ; FILING DATE: 09-JAN-1998
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/781,734
 ; FILING DATE: 10-JAN-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Einhorn, Gregory P.
 ; REGISTRATION NUMBER: 38,440
 ; REFERENCE/DOCKET NUMBER: 023070-078810US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 101:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 487 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FEATURE:
 ; NAME/KEY:
 ; LOCATION: 1..487
 ; OTHER INFORMATION: /note="RG2G deduced sequence"
 US-09-004-838-101

Query Match 2.5%; Score 107; DB 4; Length 487;
 Best Local Similarity 20.1%; Pred. No. 0.035;

Matches 89; Conservative 63; Mismatches 167; Indels 124; Gaps 21;

Oy 237 RHLYVGCGPGLVYLDLDDGMSIGHDS---DPTKSGMNTV---AGEQMPCALKFQENV 290
 Db 61 READGGKKNKFLVILDDWQFVLEDIGLSPLPNKGVNFKVLLTSDSHVCTLMGAEAN- 119
 Oy 291 KFRDVYVNEKATGPRAQOKGMKAFI-----DELKGEFKTVEHYVYVHALGYYGGLRPOVP 345
 Db 120 ---SLINKVLKDVSGQSLFROFANKAGDDDLDPAPFN-----GINDSIASRCQ 164
 Oy 346 GLPEARVIOPLVSPGLQMTMEDLAVDKIVLKRKVLGPPEKAEMEYGLHAHLEKVGIDCV 405
 Db 165 GLPIAIKRTIALSLKGRSKSAMDVALSRLNHRKIG--SEEVVREVF-----KISYDNL 214
 Oy 406 KIDVYHLEMLC-----ED-----YGRVVDL---AKATYKAMT-----KSN 439

Db 215 QDEVTKSIFLLCALFPEDDIPTEELRYVGWGLKLEIAKTIAREARNRLNCTERLETN 274
 Oy 440 KHFKN--GVIASMEHCNDF---MFLGTEAISLGRVGDDPFWCTDPSGDPNGFTWLOGCH 494
 Db 275 LFFGSDIGCVKMHVDYRDFVLHIFSEVGHASIVNHG-----NVSMLEENHS 322
 Oy 495 VH-----CANDSLMGNFIHPDWMFQSTHPCAFHAASRAISGGPIYSDSVGKH- 545
 Db 323 IYSCRISLTCKGMSQEPKDLKFPNLSILKLMH-----GDKSLG---FPEPFYCKME 371
 Oy 546 ----NFDLKKLVLPDG-----SLKSEYALPTROCLFEDPLNGETM----- 585
 Db 372 KVOVISYDKLMPPLPSSLECSSTNVVRVHLHYCSLMPFDCSSIGNLLNMEVLSFANSME 431
 Oy 586 ---LKIWNLNKFGVIGAFNCG 605
 Db 432 WLPSTIGNLKKLR-LIDLTKCKG 453

RESULT 6

US-09-004-838-47
 ; Sequence 47, Application US/09004838
 ; Patent No. 6350933
 ; GENERAL INFORMATION:
 ; APPLICANT: Michelmore, Richard W.
 ; APPLICANT: Meyers, Blake
 ; TITLE OF INVENTION: Procedures and Materials for
 ; NUMBER OF SEQUENCES: 140
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/004.838
 ; FILING DATE: 09-JAN-1998
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/781,734
 ; FILING DATE: 10-JAN-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Einhorn, Gregory P.
 ; REGISTRATION NUMBER: 38,440
 ; REFERENCE/DOCKET NUMBER: 023070-078810US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 47:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 488 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FEATURE:
 ; NAME/KEY:
 ; LOCATION: 1..488
 ; OTHER INFORMATION: /note="RG2G protein"
 US-09-004-838-47

Query Match 2.5%; Score 107; DB 4; Length 488;
 Best Local Similarity 20.1%; Pred. No. 0.035;

Matches 89; Conservative 63; Mismatches 167; Indels 124; Gaps 21;

2 IP: 332606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/291,368
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/597,607
FILING DATE: 15-OCT-90
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA50.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904)375-8100
TELEFAX: (904)372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1174 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: aizawai
INDIVIDUAL ISOLATE: PS81A2
IMMEDIATE SOURCE:
LIBRARY: lambdagem - 11 (tm) Library of August Sick
CLONE: 81A2

[illegible][illegible]

RESULT 9
US-08-962-190-2

; Sequence 2, Application US/08962190
; Patent No. 5985267

```

; GENERAL INFORMATION:
;
; APPLICANT: Payne, Jewel M.

```

APPLICANT: Sick, August J.
TITLE OF INVENTION: NO. 5985267e1 Bacillus thurthiensis Isolates

; TITLE OF INVENTION: Active Against Lepidopteran Pests
 ; NUMBER OF SEQUENCES: 27
 ;

;;
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville

STATE: FL
COUNTRY: US

ZIP: 32606
COMPUTER READA

MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible

```

; COMPILE: IBM PC COMPATIBLE
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatonTrn Release #1.0

```

```

SOFTWARE:  falconlin release #1.0/  version #1.2
CURRENT APPLICATION DATA:
ADDITIONAL INFORMATION:  uc 008/0623 100

```

APPLCATION NUMBER: 05/06/302,190
FILING DATE: 05/06/2005
REGISTRATION: 05/06/2005

CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/291,
FILING DATE: 01/03/2007

APPLICATION NUMBER: 07/597,607
FILING DATE: 15-OCT-90

ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA

TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100

TELEFAX: (904) 372-5800
; INFORMATION FOR SEO ID NO: 2

SEQUENCE CHARACTERISTICS:
LENGTH: 1174 amino acids

```

TYPE: amino acid
STRANDEDNESS: single

```

MOLECULE TYPE: protein

ANTI-SENSE: NO

ORIGINAL SOURCE: ORGANISM: Bacillus thuringiensis

ORGANISM: BACILLUS CILII
STRAIN: aizawai
IDENTIFYING ISOLATE: DC8311

INDIVIDUAL ISOLATE: PS61
IMMEDIATE SOURCE: 31

```

; LIBRARY: LambdaGem - 11
; CLONE: 81A2
;

```

US-08-962-190-2

Query Match	2.58;
Best Local Similarity	19.68;

Matches 111; Conservative 8

QY 336 YWGGLR-----PQVPGLPE-
||| ||| :

```
Query Match      2.5%:   Score 107; DB 2; Length 1174;  
Best Local Similarity 19.6%; Pred. No. 0.17/  
Matches 111; Conservative 82; Mismatches 181; Indels 192; Gaps 29  
  
Oy 336 YWGCIR-----PVGGLPE-----ARVQPVISPLQLMTMELA-----VDK 372  
||| ||| : | | | ||| : |
```

```

1      RESULT 9
2      US-08-962-190-2
3      : Sequence 2, Application US/08962190
4      : Patent No. 5985267
5      : GENERAL INFORMATION:
6      : APPLICANT: Payne, Jewel M.
7      : APPLICANT: Sick, August J.
8      : TITLE OF INVENTION: No. 5985267el Bacillus thuringiensis Isolates
9      : TITLE OF INVENTION: Active Against Lepidopteran Pests
10     : NUMBER OF SEQUENCES: 27
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: Saliwanchik & Saliwanchik
13     : STREET: 2421 N.W. 41st Street, Suite A-1
14     : CITY: Gainesville
15     : STATE: FL
16     : COUNTRY: US
17     : ZIP: 32606
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Floppy disk
20     : COMPUTER: IBM PC compatible
21     : OPERATING SYSTEM: PC-DOS/MS-DOS
22     : SOFTWARE: Patent In Release #1.0, Version #1.25
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/08/962,190
25     : FILING DATE:
26     : CLASSIFICATION: 424
27     : PRIOR APPLICATION DATA:
28     : APPLICATION NUMBER: US/08/291,368
29     : FILING DATE:
30     : APPLICATION NUMBER: 07/597,607
31     : FILING DATE: 15-OCT-90
32     : ATTORNEY/AGENT INFORMATION:
33     : NAME: Saliwanchik, David R.
34     : REGISTRATION NUMBER: 31,794
35     : REFERENCE/DOCKET NUMBER: MA50.C1
36     : TELECOMMUNICATION INFORMATION:
37     : TELEPHONE: (904)375-8100
38     : TELEFAX: (904)372-5800
39     : INFORMATION FOR SEQ ID NO: 2:
40     : SEQUENCE CHARACTERISTICS:
41     : LENGTH: 1174 amino acids
42     : TYPE: amino acid
43     : STRANDEDNESS: single
44     : TOPOLOGY: linear
45     : MOLECULE TYPE: protein
46     : HYPOTHEetical: YES
47     : ANTI-SENSE: NO
48     : ORIGINAL SOURCE:
49     : ORGANISM: Bacillus thuringiensis
50     : STRAIN: aizawai
51     : INDIVIDUAL ISOLATE: PS81A2
52     : IMMEDIATE SOURCE:
53     : LIBRARY: Lambdagem - 11 (tm) Library of August Sick
54     : CLONE: 81A2
55     : US-08-962-190-2

```

```

Db 314 YWAGHRTVSHFTGSSQVSSIPQYGITANAEPSTRTIAPSTPGLNLFYRTLSDPFFRRSDN 373
Qy 373 I-----VLKRVGLVPEPKAEEMT--EGLHAHLEKVGIDGK--IDYIHLEMLCEDYG 421
Db 374 IMPTLGIVNVOGVGFIOQNGEVLRRRGTYDSDELPIDGENSLVGSYSHRL----- 425
Qy 422 GRVDLAKAYVKAMTKSI-----NKHFGKGVIASMEHCNDF-MFLGTEAI-SLG 468
Db 426 SHVTLTRSLYNTNITSLPTFWTHHSATDRNLIYDPVITQIPLVKSFLTSGTSVVRGPG 485
Qy 469 RVGDDEFWCTDPSGD-----PNCFFMLQGGCHM 494
Db 486 FTGGDIIRTNVNGVNLMSLNFSTLQRYRVRYAASQTMWVRVNGSTTFDQGFPS 545
Qy 495 VHCANDSLMGNFIHPDMDFOSTHPCAFHAASRAISGPIYVSDYSGKHNFDLKLTV 554
Db 546 TMSANGSLTSGSFRAEPVPGIST-----SGSQTAG-ISTINNPGRCQTFHLDRIEF 595
Qy 555 LPDGSILRSEY--YALPTRDCLFEDPLHNGETMLKI---WNLKKFTGVIGAFNCGGG 607
Db 596 IPVDATFEAEYDLERAQAANSFLTSS--SNOIELKTDVTDYHIDQVSNLV--DCLSD 649
Qy 608 WCRETNRNOCES-QYSKRVTSKTN-----PKDIEMHSG-----EN 641
Db 650 FCLDEKRELSEKVKHAKRLSDERNLLODPNFRGINRQPDGRGSGTDITIQGGDVFKEN 709
Qy 642 PISIEGV--KTFALVLYQAKKLILSK-----PSQDDIALDP--EFELIT 683
Db 710 YVTLPGTDECPYLYO--KIDESKLAAYNRQYLRGYIEDSODLEIYLIRYNAKHETVN 767
Qy 684 VSPVTKLIQTSLHFAPIGLVNLMTSGAIQSVYDDLS-----SVEIG 727
Db 768 VPGTSLPLSLVE--SPIRGCEPNR--CVPHELEWNPDLDCSCRODEKCAHSHHPSLID 824
Qy 728 VKGCGEMRVFASKRPACRIDGEDVG 753
Db 825 V-GCTDLO-----EDLG 835

```

```

RESULT 10
Sequence 2, Application PC/TUS9510310
GENERAL INFORMATION:
APPLICANT: MYCOGEN CORPORATION 5501 Oberlin Drive
APPLICANT: STREET ADDRESS:
APPLICANT: CITY: San Diego
APPLICANT: STATE/PROVINCE: California
APPLICANT: COUNTRY: US
APPLICANT: POSTAL CODE/ZIP: 92121
APPLICANT: PHONE NUMBER: (619) 453-8030
APPLICANT: FAX NUMBER: (619) 453-6991
TITLE OF INVENTION: Protein Toxins Active Against Lepidopteran Pests
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESS: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10310
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/597,607
FILING DATE: 15-OCT-90
CLASSIFICATION:

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA50.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904)375-8100
TELEFAX: (904)372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1174 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: aizawai
INDIVIDUAL ISOLATE: PS81A2
IMMEDIATE SOURCE:
LIBRARY: LambdaGem - 11 (tm) Library of August Sick
CLONE: 81A2
PCT-US95-10310-2

```

Query Match 2.5%; Score 107; DB 5; Length 1174;

Best local similarity 19.6%; Pred. No. 0.17; Mismatches 181; Indels 192; Gaps 29;

```

Matches 111: Conservative 82; Mismatches 181; Indels 192; Gaps 29;
Qy 336 YWGLR-----POVGLPE-----ARVIOPLYSPGLQMTMEDLA-----YDK 372
Db 314 YWAGHRTVSHFTGSSQVSSIPQYGITANAEPSTRTIAPSTPGLNLFYRTLSDPFFRRSDN 373
Qy 373 I-----VLKRVGLVPEPKAEEMT--EGLHAHLEKVGIDGK--IDYIHLEMLCEDYG 421
Db 374 IMPTLGIVNVOGVGFIOQNGEVLRRRGTYDSDELPIDGENSLVGSYSHRL----- 425
Qy 422 GRVDLAKAYVKAMTKSI-----NKHFGKGVIASMEHCNDF-MFLGTEAI-SLG 468
Db 426 SHVTLTRSLYNTNITSLPTFWTHHSATDRNLIYDPVITQIPLVKSFLTSGTSVVRGPG 485
Qy 469 RVGDDEFWCTDPSGD-----PNCFFMLQGGCHM 494
Db 486 FTGGDIIRTNVNGVNLMSLNFSTLQRYRVRYAASQTMWVRVNGSTTFDQGFPS 545
Qy 495 VHCANDSLMGNFIHPDMDFOSTHPCAFHAASRAISGPIYVSDYSGKHNFDLKLTV 554
Db 546 TMSANGSLTSGSFRAEPVPGIST-----SGSQTAG-ISTINNPGRCQTFHLDRIEF 595
Qy 555 LPDGSILRSEY--YALPTRDCLFEDPLHNGETMLKI---WNLKKFTGVIGAFNCGGG 607
Db 596 IPVDATFEAEYDLERAQAANSFLTSS--SNOIELKTDVTDYHIDQVSNLV--DCLSD 649
Qy 608 WCRETNRNOCES-QYSKRVTSKTN-----PKDIEMHSG-----EN 641
Db 650 FCLDEKRELSEKVKHAKRLSDERNLLODPNFRGINRQPDGRGSGTDITIQGGDVFKEN 709
Qy 642 PISIEGV--KTFALVLYQAKKLILSK-----PSQDDIALDP--EFELIT 683
Db 710 YVTLPGTDECPYLYO--KIDESKLAAYNRQYLRGYIEDSODLEIYLIRYNAKHETVN 767
Qy 684 VSPVTKLIQTSLHFAPIGLVNLMTSGAIQSVYDDLS-----SVEIG 727
Db 768 VPGTSLPLSLVE--SPIRGCEPNR--CVPHELEWNPDLDCSCRODEKCAHSHHPSLID 824
Qy 728 VKGCGEMRVFASKRPACRIDGEDVG 753
Db 825 V-GCTDLO-----EDLG 835

```

```

RESULT 11
5164180-4
Patent No. 5164180

```

```

; APPLICANT: Payne, Jewel, Sick, August J.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS ISOLATES ACTIVE
; AGAINST LEPIDOPTERAN PESTS
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/451,389
; FILING DATE: 14-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 353,860
; FILING DATE: 18-MAY-1989
; SEQ ID NO: 4
; LENGTH: 1174
5164180-4

Query Match          2.5%; Score 107; DB 6; Length 1174;
Best Local Similarity 19.6%; Pred. No. 0.17;
Matches 111; Conservative 82; Mismatches 181; Indels 192; Gaps 29;

QY 336 YWGLR-----PQVPLPE-----ARVIOPLSPGLQMTMEDLA-----VDK 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 314 YWAGHVTSHFTGSSQVISPQGITANAPERTIAPSTFPGINLFRSLDFFRRSDN 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 373 I-----VLHVGVLPPEKAEEMV--EGLHALEKVGIDYK--IDVHLLMLECEDYG 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 374 IMPTLGINVGQVGFLOPNNGEVLNRRGTVDSDLDPLDGENSLGVSHRL----- 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 422 GVDLAKAYKATKSI-----NKHFGNGVIAEMHCNDF-MELGTETAI-SLG 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 426 SHVTLRLSLYNINITSLPFWVTHSATDRNLIYPIVITQIPVKFSFSLTSGSVVRPG 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 RVGDWFCTDPSGD-----PNGTFLWLGCHM 494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 486 FTGGDIIRTNVNGVLSMSLNFNTSLQRYRVRRVRYAASQTMVRYNVGSGITFDGFPFS 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 495 VHCANDSLMGNFIHPDMFQSTHPCAFHAASRAISGPIYDVSVCCKHNDLKLKIV 554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 546 TMSANGSLISQSFRAEFVVGIST-----SGSQTAG-ISISNPNRGQTHLDRIEF 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 555 LPDGSILRSEY--VALPTRCLFEDPLHNGETMLKTI---WNLNKFTGVIGAFNCGGG 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 IPVDATFEAEYDLERAKAVNSLFTS--SNQELKTYDTYHIDVSNLV--DCLSDPE 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 608 WCRFRNRNOCFS-QYSKRYTSKTN-----PQDIEMHSG-----EN 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 650 FCLDERRELSEKVKAKRSLDERNLQDPNFRGINQPDGWSGSDITITGCGDVEKEN 709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 642 PLSIEGV--KTFALYLYOKKILSK-----PSODLIAIDPF--EPELIT 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 710 YVTLPGTPEFCPTTYLXO--KIDESKLAVNRQLGCIYEDSODLEILIRYNAKHETVA 767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 684 VSPVTKLIQTSLHFAPIGLVNMLNTSGAIOVDYDDL$-----SVEIG 727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 768 VPGTGLMPLSYE-SPIGRCGEPRN--CVPHLEWMPDLDCSCRGKCAHSHHFSFLDID 824
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 728 VKGCGEMRVFASKKPRACRIDGEDVG 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 825 V-GCTDLO-----EDLG 835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-08-846-234-2
; Sequence 2, Application US/0846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBILON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P. C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR

```

```

; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBILON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
US-08-846-234-2

Query Match          2.5%; Score 105; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 PVSVCGEVGFDASEPDSRH 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 PVSVCGEVGFDASEPDSRH 19

RESULT 13
US-09-819-993-2
; Sequence 2, Application US/0981993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-09-819-993-2

Query Match          2.4%; Score 102; DB 4; Length 478;
Best Local Similarity 18.5%; Pred. No. 0.11;
Matches 77; Conservative 54; Mismatches 95; Indels 190; Gaps 19;

QY 378 VGLVPE-----KAF-EMTEGLHAHLEKVGIDGKID-----VIHL 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17 VGLVLEIYFPSSQYDQAELEKYGVDAGKTYTIGLQAKGFCCTDRBDINSICMTVVQNL 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 413 LEM--LCEDYGVADLAKAYKATKSI-----KNGVIAEMHCNDFMPLGTET 464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77 MERNNLSTYDIGLVEGTETIIDKSKSVKTNLMQLEESGNTIDIGIDITN-ACYGATA 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 465 ISLGRVDDFWCTDPSGDPNGTFMLOGCHVHCANDSLMGNFIHPDMDFQSTHPCAAF 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 VF-----NAVNWIES-----SSWDGLRGTTH----- 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 525 HAASRAISGPIYVSVGCKHNFDDLKLVLPDGSILRSEYVALPTRCLFEDPLHNGET 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 157 -----QHAIDEYK-----PD-----MLSEY-----PIVDGKL 178
OY 585 MLKIMNLKFTGVIGAFNCGGNGCERTRRNCFCFSYKSRVSKTNPKDIEHWSGENPIS 644
Db 179 SIO-----CYLSALDRCYSYCKKHA-----QW-----202
OY 645 IEGVTFALYVQAKKLISKPSODDIALDPFEFLITVSPVKLTQSLHFAPIGLVN 704
Db 203 -----QKGNOKDFTLNDGFG-MIFHSPYCKLVQKSL--AMRLND 240
OY 705 MLNTSGAIOVDYDDDLSSVEIGVKGCEMRVFASKRPACRIDEDVGFKYDQD 760
Db 241 FLN-----DQNRDKNSIYSGLFAFGDKL-----EDTYFDRDVER 275

RESULT 14
US-09-819-993-5
Sequence 5, Application US/09819993
Patent No. 6436692
GENERAL INFORMATION:
APPLICANT: GONG, Fangcheng et al.
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
FILE REFERENCE: CLO01195
CURRENT APPLICATION NUMBER: US/09/819,993
CURRENT FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 5
LENGTH: 518
TYPE: PRT
ORGANISM: Human
US-09-819-993-5

Query Match 2.4%; Score 100; DB 4; Length 518;
Best Local Similarity 19.0%; Pred. No. 0.21;
Matches 82; Conservative 59; Mismatches 111; Indels 180; Gaps 20;

OY 378 VGLVPE-----KAE-EMVEGLHAHLEKVGIDGVKID-----VHIL 412
Db 17 VGIVALEIYFPQYVDQALEKYGVDGAKYITIGLOAKMGCTDREDINSICMTVQNL 76
OY 413 LEM--LCEDYGRVDLAKAYYKAMTKSINKHF-----KNGCVIASMEHNDPFLGTEA 464
Db 77 MERNNLSTDCIGRLEVGFTTIDKSKSVKTNLMQLFEESGNTDIEGIDITN--ACYGCTAA 135
OY 465 ISLGRVGDGFWCTDPSGDPNGTFMLQGCIMVHCANDSLMGNFIHPDMDFQSTHPCAA 524
Db 136 VF-----NAVMWIE-----SSWDGRY-----ALVAGDIAY 163
OY 525 HAASRAISGPIYSDSVG-----KHNFDLKLKLVLPDGSILRSEYVAL 568
Db 164 ATGNARPTGCVAVALLIGPNAPLIFERGLRGTNHQHAIDFK-----PD--MLSEY--- 213
OY 569 PTRDCLFEDPLNGETMLKIMLNKFTGVIGAFNCGGNGCERTRRNCFCFSYKSRVSK 628
Db 214 -----PIVDGKLSIO-----CYLSALDRCYSYCKKHA-----242
OY 629 TNPKDIEMHSGENPISIEGVKTFALYVQAKKLISKPSODDIALDPFEFLITVSPVT 688
Db 243 -----QW-----QKANDNDFTLNDGFG-MIFHSPYC 268
OY 689 KLIOTSLHFAPIGLVNMLNTSGAIOVDYDDDLSSVEIGVKGCEMRVFASKRPACRID 748
Db 269 KLVQKSL--ARMLNDFLN-----DQNRDKNSIYSGLKAFGDKL-----306
OY 749 GEDVGFKYDQD 760
Db 307 -EDTYFDRDVER 317

RESULT 15
US-08-305-505-6
Sequence 6, Application US/08305505
Patent No. 5668001
GENERAL INFORMATION:
APPLICANT: Maziorko, Henry M.
TITLE OF INVENTION: 3-HYDROXY-3-METHYLCULFARYL-COA
TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
TITLE OF INVENTION: STABILITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quaries & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/305,505
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/072,040
FILING DATE: 02 JUNE 1993
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 65-053-9083-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5709
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 520 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-305-505-6

Query Match 2.4%; Score 100; DB 1; Length 520;
Best Local Similarity 19.0%; Pred. No. 0.22;
Matches 82; Conservative 59; Mismatches 111; Indels 180; Gaps 20;

OY 378 VGLVPE-----KAE-EMVEGLHAHLEKVGIDGVKID-----VHIL 412
Db 17 VGIVALEIYFPQYVDQALEKYGVDGAKYITIGLOAKMGCTDREDINSICMTVQNL 76
OY 413 LEM--LCEDYGRVDLAKAYYKAMTKSINKHF-----KNGCVIASMEHNDPFLGTEA 464
Db 77 MERNNLSTDCIGRLEVGFTTIDKSKSVKTNLMQLFEESGNTDIEGIDITN--ACYGCTAA 135
OY 465 ISLGRVGDGFWCTDPSGDPNGTFMLQGCIMVHCANDSLMGNFIHPDMDFQSTHPCAA 524
Db 136 VF-----NAVMWIE-----SSWDGRY-----ALVAGDIAY 163
OY 525 HAASRAISGPIYSDSVG-----KHNFDLKLKLVLPDGSILRSEYVAL 568
Db 164 ATGNARPTGCVAVALLIGPNAPLIFERGLRGTNHQHAIDFK-----PD--MLSEY--- 213
OY 569 PTRDCLFEDPLNGETMLKIMLNKFTGVIGAFNCGGNGCERTRRNCFCFSYKSRVSK 628
Db 214 -----PIVDGKLSIO-----CYLSALDRCYSYCKKHA-----242
OY 629 TNPKDIEMHSGENPISIEGVKTFALYVQAKKLISKPSODDIALDPFEFLITVSPVT 688
Db 243 -----QW-----QKANDNDFTLNDGFG-MIFHSPYC 268


```
OY 301 TGPRAQKMKAFIDE-----LKGEFTVEHYVW-HAL-CGYNGGLRPQVPLPEAR 351
DB 243 ---NMSSSESDSSCDTSETCVPLEGEKSAEKKRVRFOHLSCLINDRNKLLPEF----- 295
OY 352 VIQPVSPGLQMTME---DLAVKIVLHKV-----GLVPEKAEMEYEGHHALEK 399
DB 296 VROFSIDRCCEMTPKTPGDLAWN--FLMKYQALDSTARSDILRPEVAGEENELPAGITEK 333
OY 400 VGIDGKVIDIHLE-----MLCEDYGGRVDLAKAYYKAM-----TKSINKHFKNG 446
DB 354 LGIGDQP--TIHPLDVLACMLCADSSLRREVSNMYQCOFALPRLPDAENKNKLLMWG 411
OY 447 VIASMHCHN-----DPMFLGTEAISLGRVGDDEFWCTDPSGDP 483
DB 412 AMKDLQPSQOSSGGLRETFELGLTKMPVISFVRLGRCSFSKSRIVNTL--LSSSQOKP 470
OY 484 NGTFWLOGCH---MVHCANDSLMGNFIHPDMDPOSTH---PCAFAHNR----- 529
DB 471 YPFLHODLSVPVLPQOISGLVEVTWCPPDKLLKESPHAFQKPVAVANLRGDLSEFWIO 530
OY 530 ---AISGPIYVSDSVGKHNFEDLLKLVLPDGSILRSEYALPTRDCLFEDPLHNGET 584
DB 531 FGLVEVSSGLFEFTDCEKEWDL--MFLGSDTIERCFILSPQ----- 574
OY 585 MKTIMNLNKTGVTGAFNCOGGGWCETRRNOCFSQYSKRWYSKTPNKDIEM--HSGEN 641
DB 575 -----AKESEBAQIFORILKIKPSQL-----LFWEAEEAGDR 606
OY 642 PISIEGVTFEALYQAKLLISKP-----SODLIALDPEFEEL---ITVSP 686
DB 607 RKTMEA-----LQALDQEWSSPLRCVSLREMASLARLGIQVDO--DEVTQDIOVSP 658
OY 687 VT 688
DB 659 TT 660

RESULT 2
US-10-193-295-2
; Sequence 2, Application US/10193295
; Patent No. US20020173018A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CLO01195DIY
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-10-193-295-2

Query Match
Best Local Similarity 18.5%; Pred. No. 0.63; Length 478;
Matches 77; Conservative 54; Mismatches 95; Indels 190; Gaps 19;

OY 378 VGLVPE-----KAE-EMEGHHALEKVIDGVRID-----VIHL 412
DB 17 VGLVLEIIFPQSOYDQDALEKDYDACKYITIGLOAKMGCTDREDINSICMTYVONI 76
OY 413 LEM--LCEDYGGRVDLAKAYYKAMTKSINKH-----KNGVYASMEHNDPMLGTGA 464
DB 77 MERNNLSTYDPCIRLEVEGTETIIDSRSVKTNLMOQLFEESGNDIDIEDITTN--ACYGTAA 135
OY 465 ISLGRVGDDEFWCTDPSGDNFTWLOGCHMVHCANDSLMGNFIHPDMDPOSTHPCAFA 524
```

```
DB 136 VF-----NAVNWIES-----SSWDGLRGTHM----- 156
OY 525 HAASRAISGPIYVSDSVGKHNFEDLLKLVLPDGSILRSEYALPTRDCLFEDPLHNGET 584
DB 157 -----QHAYDFK---PD--MLSEY-----PIVDGKL 178
OY 585 MKTIMNLNKTGVTGAFNCOGGGWCETRRNOCFSQYSKRWYSKTPNKDIEMHSGENPIS 644
DB 179 SIQ-----CYLSALDRCSYCYCKIHA-----QW----- 202
OY 645 IESEVTFEALYQAKLLISKPSODLIALDPEFEELITVSPYTKLIQTSILHRAPIGLVN 704
DB 203 -----QKEGNDKPFITNDQGF--MIFHSPYCKLVQXSL--ARMNLND 240
OY 705 MLNTGAIQSDVDYDDLLSSVEIGKCGGEMRVFASKPRACRIDGSDVGFKKYDQDQ 760
DB 241 FLN-----DGNRRKNSTIYSGLFAFGVKL-----EDTYFDRDYEK 275
```

```
RESULT 3
US-10-001-851-27
; Sequence 27, Application US/10001851
; Patent No. US20020115628A1
; GENERAL INFORMATION:
; APPLICANT: MEYERS, Rachel A.
; APPLICANT: WILLIAMSON, Mark
; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1 Human Glycosyl Transfer
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 10147-5601
; CURRENT APPLICATION NUMBER: US/10/001,851
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: US 60/249,939
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-001-851-27
```

```
Query Match
Best Local Similarity 21.3%; Pred. No. 1.1; Length 626;
Matches 77; Conservative 52; Mismatches 118; Indels 115; Gaps 20;

OY 220 WDAFYLVTHPQGVIEGV-RHLVDGCGPGLVLIDGWSIGHSDPITREGMNQTVAGEO 278
DB 188 WSLVLTFTVH--SVLETPPHLE-----EVLAVDD--FSDMDHTKRPL--EYMSQ--FGKRV 237
OY 279 MPCRLLFQENKFKRDYVNRKATG-----RAGQKMKAFIDELKGERXTV----- 324
DB 238 KILRMEKREGILIRARLGAAVATGEVLYTLDSHCEGMEGMEBLDRIRKDPPTVVCPI 297
OY 325 ---EHVYVWH-----ALCGYWGGLRPQVRLPE-----ARVQVPLSPQLQMTMED 367
DB 298 DAVIDMTFEYHHSKAYFTSVGGFWGLQFNWHSIPBRDRKNRTPTIDPASP---TMAG 353
OY 368 --LAVDKIVLHKV-----LVPEKAEMEYEGHHAHL 397
DB 354 GUFSTIKYEFEKIGTIDPGRFDITMGGENLSTFRIMCGGTLEIVPESHGHVFRKRSPTK 413
OY 398 EKVAGIDVKAIDVLIHLEMLCEDYGGRVDLAKAYYKAMTKSINKHFGNGVYASMEH--- 453
DB 414 WRGVAVNLKNSIRLAELVWLDY-----KTY--YERINNQLGDFDISRKKLRD 463
OY 454 ---CNDP-----MELGTEAISLGRVGD---FWCTD-----PSCDNGRTWLOG 491
DB 464 LOCKSRKWTLDNIYPRLFVPGESVAVAGEVNSAVOPARCLDCWVGHEKRNRPVGYT--Q 520
OY 492 CH 493
DB 521 CH 522
```

```
RESULT 4
US-10-193-295-5
; Sequence 5, Application US/10193295
; Patent No. US20020173018A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEROP
; FILE REFERENCE: CLO0119501V
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Human
US-10-193-295-5

Query Match          2.4%; Score 100; DB 9; Length 518;
Best Local Similarity 19.0%; Pred. No. 1.1;
Matches 82; Conservative 59; Mismatches 11; Indels 180; Gaps 20;

OY 378 VGLVPE-----KAE-EMEGHALLERKVGIDGVKID-----VHIL 412
DB 17 VGLVLEIYFSPQYQVDALEKRYDGDVDAKTYTIGLQAKMGFCTDREDINSICMTVVOQL 76
OY 413 LEM--LCEDYGRVDLAKAYKAMTKSINKHP-----KNGVIASMEHCNDPMEFGTEA 464
DB 77 MRRNNLSYVCIRLEVGTTITIDKSKSVKTNLMQLEFEEGNDIDIGIDITN-ACYGCTA 135
OY 465 ISLGRVGDDEFWCTDPSGDGNGTFMLQGCCHVHCANDSLMGNFIHPDWMDFOSTHPCAA 524
DB 136 VF-----NAVWIE-----SSWDGRY-----ALVAGDIAY 163
OY 525 HAASRAISGGRPIYVSDSG-----KHNDLKLKLVLPDGLSLRSYVAL 568
DB 164 ATGNARPTGCGAVALLGPNAPLIERGLRGTTHQHAIDFYK---P--MLSEY--- 213
OY 569 PTRDCLFEDPLHNGETMLKIMLNKFTGIVIGAFNCGGCMGCFETRRNOCFSQYSKRVTSK 628
DB 214 -----PIYDGLKLSIO-----CISLALDRKYCYCKKIHA- 242
OY 629 TNPKDIEMHSGENPISIEGVKTFALYLYQAKKLILSKPSQDLDIALDPREFELIVSPVT 688
DB 243 -----QW-----OKEANDNDFTLNDGCF-MIFHSPLYC 268
OY 689 KLIQTSIHAPFGLVNMNTSGAIOVDYDDDLSSVEIYKCGEGEKRVASAKPRACRID 748
DB 269 KIVQKSL-ARMLNDLDFLN-----DQNRDKNSIYSLGLAKAFGDKL----- 306
OY 749 GEDVGFKYDQD 760
DB 307 -EDTYFDRDYER 317

RESULT 5
US-10-095-718-4
; Sequence 4, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Bursstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated virus Vectors Encoding Factor VIII and
; TITLE OF INVENTION: Methods of Using the Same
```

```
; FILE REFERENCE: 350522/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match          2.3%; Score 98; DB 12; Length 1431;
Best Local Similarity 19.0%; Pred. No. 8.2; Indels 316; Gaps 33;
Matches 133; Conservative 65; Mismatches 186; Indels 316; Gaps 33;

OY 11 NVVSFDGLNDSSPFAIDSDPTVNGH-----SFLSDVENIVASPSPTSIDKS----- 60
DB 369 DVVSED--DDSSSPFIQIRS--VAKKHPTWVHYIAAEEEDMDYAPSGFTPDNRSHKILY 424
OY 61 ---PVSVGCFVGFAPDASEPDSRHVVSICKLDIRPMSIFRFKVVMTTHVGNNGDL--- 113
DB 425 LNNGPQRIQ-----KKYKVRFAA-VTDEFPTREALIOYESGILGPL 465
OY 114 ---ESETQVILEKSDSGRPYFLLPIYEGPRTSIOBPDDDFVDVYCVSSGSKVYDASF 170
DB 466 LYGEVGDTLILIFKQASPR-----NIYPGIVYVVP----- 498
OY 171 RSMVLVHAGDDPFALYKEAMKIVRTHLGTFLEERTPGIVDKFGQCTW-----DA 222
DB 499 -----LHTGRLPKV-----KILKDPIL-----PGEIFKRYKYTVYEDGPRKSP 539
OY 223 FYLVNHPQGVIEGRHLVDGCRPGLVLIDGMQSIGHSDPITKGMQVAGDEQMPCR 282
DB 540 RCLTRYSSFINLERDLASGLIGPLLICYESVDQG-----NQMSDKRNVL 588
OY 283 LKFOENYKFRDYVNPKATGPRAGQGMKAFIDELKGEFTVEHYVWHAALCGYMGGLRP 342
DB 589 FSVFEDENRSM--YLQ-----ENMQRF----- 607
OY 343 QVPGLPPEARVIOVPSGLQMTMEDLAVDKIVLHKVGLVPRPEKAEEMYGLHAHLEKVI 402
DB 608 ---LPNADVQRP-HDPERQLSN-----IMH-----SI 630
OY 403 DGVKIDVILHLEMLCE-----DYGGRVDLAKAYKAMTKSINKH-----FK 443
DB 631 NGYVFDNLQLSVCLHEVAWYILTSGAQOTDPLSVFPGYT---FKHKMYEDTLTLFPSPS 687
OY 444 GNGVLAAME-----HCNDPMLGTEAT---SLGAVGDDFW----- 475
DB 688 GETVFMSEMENPGIWLVLGCHNSDFRNGMTALLKVSCHNRNIDYEDYEDIPPTLINEN 747
OY 476 -----CTDPSG-----DPNGTFMLQGC-----HMVHCANDSLAM 504
DB 748 NVIKPRFSQNSRHSPTKQOLKMKREDFDIYGVENQGLRSFQKTRHYFTAAVERLNL- 806
OY 505 GNFHPDMDFOSTHPCAAFAHAASRAISGPIYVSDSVGKHNHFDLLKLV---LPDGLSL 561
DB 807 -----DYGMSRSPH-----ILNRNAGS-----DYGOCKKVVFPQEPFDGS-- 841
OY 562 RSEYIALPTRDCLFEDPLHNGETMLKIMLNKFTGIVIGAF 601
DB 842 -----FTQPLRYGE-----LNEHLGILGPY 861

RESULT 6
US-09-978-522-1
; Sequence 1, Application US/09978522
; Publication No. US2003003627A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Descenzo, Richard
; APPLICANT: Ireland, Nancy
; TITLE OF INVENTION: Lipoygenase Genes From Vitis Vinifera
; FILE REFERENCE: 29520/37890
; CURRENT APPLICATION NUMBER: US/09/978, 522
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/241,220
; PRIOR FILING DATE: 2000-10-16
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Vitis LOX 1
US-09-978-522-1

Query Match      2.3%   Score 96.5:  DB 9;   Length 862;
Best Local Similarity 19.1%;  Fred. No.5.1;
Matches 109;  Conservative 76;  Mismatches 203;  Indels 183;  Gaps 29.

QY  133  FLPLPVEGPFRTSIOPGC-DDFVVC-VESGSSKVVADSFMSRLYLHGDDPFAIVKEM 190
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  286  FLPLPEFA-LCDIRPNEDSFQVDLDLYEGSIKPEBGLDKI--KDNILEMKELEY 340
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  191  KIVRHLGTFLLLEKTPPGIV--DFGKCTWDATFYLVHPOGVLEGVRLHYDGCPCPEL 248
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  341  RLDGEHLKFFPM--POYIKEDKSAMRDEEF--AREMLAGINPVYI 382
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  249  VLIDP-----GMSIGHSDSPITREKGNQ-----TYAGEQMPCKLKFQENYKFRD- 295
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  383  RLLOEPPKSKLDPEVYNGNOSTITKEHLENHLDLTTINEAMEKKRLPLLDHHDVMPYL 442
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  296  -VNPKATGPRAQK-----GKKAFLDELK-----GEFTVEHYV-----V 329
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  443  RRIINTSTKYASRFLLEFLKDGDTLKPRLAIELSLPHPNMGDKGAVNKKVTPAEDVEGSI 502
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  330  W-----HALGQYW-----GGLRPOVPEL-PEARLIOPV--LSPGLQMTMED 367
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  503  WOLAKATYAANDSGTHQLLSHMLNTHAAIEPVIATNRQLSVLHPILHKLHPHFKDTMTI 562
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  368  LAVDKIVLHKVG-----LVPEKAEM-----YEGIAHLEKVGIDGVKIDVI 410
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  563  NALAKOILINAGGVVESTVFPSKYAMEMSVYKDWMLTEQALPADLLKRGMAVEDSEAP 622
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  411  HLEMLACDYGGRVD-----LAKAYK-----AMTSINKHKRG 444
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  623  HGLRLLIDDPYAVDGLIWSALETWVKYECSEFYKTKIDEMVOKDSELQSMWKEVRE--EG 680
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  445  NG-----VIASMEHCNDPMFLGT--PAISLGR-----VEDDFWC 476
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  681  HODKDEBWMWRKMRVYKELIENCITIIIVASALHAAYVFGQPYAGLYLPNRTISRPPMP 740
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  477  TDPS-----GDPNGTF-----WLQGCWHVHC-----ANDSLMGNFIHPMDMFO 516
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  741  ECGPEVEEELKSNPKAKLKTITTAOLQTLGLISTLEVLRSHSSDEVYLGQORDTPWTL-- 798
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY  517  STHPCAFAHSAASRAISGGPIYSDSVGKHN 547
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB  799  DTPPLKAEFKGKRLADIEKMTIDRNGNERF 829
      ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 7
US-10-193-295-4
; Sequence 4, Application US/10193295
; Patent No. US20020173018A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CU001195DIY
; CURRENT APPLICATION NUMBER: US/10/0193, 295
; CURRENT FILING DATE: 2002-07-12

```

```

PRIOR APPLICATION NUMBER: 08/819,993
PRIOR FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 520
TYPE: PRT
ORGANISM: Human
US-10-193-295-4

Query Match      2.3%, Score 96; DB 9; Length 520;
Best Local Similarity 18.8%, Pred. NO. 2.6;
Matches 81; Conservative 59; Mismatches 112; Indels 180; Gaps 20;

QY 378 VGLVPE-----KAE-EMEGTNAHLEKVGIDGKID-----VITL 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17 VGIVALEIPIFSQYDQAELEKYDSVDGKTYTIGLGQAKMGECTREDINSLCTMYQNL 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 413 LEM-LCEDYGGRYDLAAAYKAMTKSINKHF-----KNGVIASMEHCNDFMFLGTEA 464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 77 MERNNLSDYDCIGRLEVGTEITIIDKSKSVKTNLMQLFEESGNTDIEDITTN-ACYGTA 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 465 ISLGNVGDDFNCTDSSGPNCTFMVIGCGHMHCAADSLMNGFIHPDMDPQSTHPCAA 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 VF-----NAVNMLE-----SSSMWGRY-----ALVYACDIIVY 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 525 HAASRAISGPIYVSDSYG-----KHFEDLTKLVLDPDGLIRSEYYAL 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 164 ATGNARPIRGVAVALLIGPNAPLIFERGLRGTTHQHAYDFYK---PD---MISEY--- 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 569 PTRDCLFEDPLHNGETMTAKIMNKNKFTVIGAFNCGGCGWCKEYTRNOCFSQYSKRVTSK 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 214 -----PIVDGKLSTQ-----CYLSALDRCYSYCKKIIHA- 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 629 TNPKDIEWHSENPISIEGVKTFALYLYQAKKLLSKPSODIDIALDPFEFELITVSPVT 688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 -----QW-----QKRGNDKDTLNDPFGF-MIFHSPYC 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 689 KLIQTSLHFADIGLVNMLNTSGAIOVDYDDDLSSVEIIGVKGCGEMRYFASKKRPACRID 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 269 KLYVKSL--ARMLLNDPLN-----DQNRDKNSIYGLAEFGDKL----- 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 749 GEDYGFKYDQD 760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 307 -EDTYFDRDYER 317

RESULT 8
US-10-121-032-28
Sequence 28, Application US/10121032
Patent No. US20020155550A1.
GENERAL INFORMATION:
APPLICANT: Bylina, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER-READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/121,032
FILING DATE: 09-Apr-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION NUMBER: US/09/134,078

```

FILING DATE: 13-AUG-1998
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hall, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 772 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-121-032-28

Query Match 2.3%; Score 96; DB 9; Length 772;
Best Local Similarity 20.6%; Pred. No. 4.8;
Matches 138; Conservative 64; Mismatches 166; Indels 302; Gaps 42;

QY 97 FKVM-----WTHHWGRNGDLESETOIYILEKSDG-----RPVFLPIVE--G 140
DB 165 FHWSPVSKVWVLLFKNEDETE-PRQVNMEEKGNGWEAVEGDDGVLYOLENYG 223
QY 141 PRTSIOPDDDFVDVVCVSSGSKVVDASFRLMLYHAGDDPFALV----- 187
DB 224 KIRTYDP-----YSKAV-----YANNQSAVVNLTARTNEGENDR 260
QY 188 -----EAMKIVRTHLGTFRLLLEKTPPGIVDKGWCMTDAFYL-----TVHPOGVIE 234
DB 261 GKRIEGEYDAIYIEIHADITGLENS--GVKNK-----GLYIGLIEENTKPGCVTT 310
QY 235 GVRHLVDGCGPRGLVLIDCGMQSIGHSDSPITKESMNTVAGEQMCRLLKROENKFRD 294
DB 311 GLSHLVELG-----VTH-----VHILPFEDFY----- 332
QY 295 YVNPKATGPRAGOKMKAFIDELKGEFTVEHYVWNLGCVWGLRPQVGLPEARVIO 354
DB 333 -----TG-----DELDRKF--EKYYNM-----GYDPLFMVPEGRYST 363
QY 355 PVLSPLOMTMEDLAVDKIVLHKVGLVPRPEKAEMTEGLHAHLEKVGIDGVKIDVIH--- 411
DB 364 DPKNPHTRI-----REYKENYKALHKH---GI-GVIMDVVFPHT 398
QY 412 --LLEMLCED-----YGVHVDLAKAVYKAMTSSINKHFKNGVYASME----- 452
DB 399 YGIGELSAFDQVPIYFYIDTGAIV-----LNESGCGN-VIASERPMKRKEIVDTVT 450
QY 453 -----HCNDFMF--LG--TEAISTLGVGDDFWCTDPS--GDPNGTFW-----L 489
DB 451 YWVKENHIDGFRDDKGLDKKTMLEVERALKIDPTIILYGEPMWG--WGAPIRFGKSUV 509
QY 490 QGCHWYHCAHDSLMNGNFIHPDMFQSTHPCAAFIHAASRA--ISGGPIYVSDS-----Y 542
DB 510 AGTH--VAAFNDE-----FRDAIRGSVVFNPISVKGFMWG--YGETKIKIRGVV 553
QY 543 GKHNFDLLKLVLPDGSILRSEYVALPTRDCLFEDPLHNGETM-----LKITV-- 589
DB 554 GSIINY-----DGKLKLS--FALDPRETTIYVACHCHNHILMDKNYLAARADKKKEVT 602
QY 590 -----NLNKFYGVIGAFNCG-----GC--WCRETRRNOCFQSYSKRVTSKTNPKDIEWH 637
DB 603 EELKNAQKLAGAI--LLTSGQVPEFLHGQDFCRTTNFND-----N 641
QY 638 SGENPISIEG 647
DB 642 SYNAPISING 651

RESULT 9
US-09-771-161A-173
Sequence 173, Application US/09771161A
Patent No. US20020110811A1
GENERAL INFORMATION:
APPLICANT: LEVINE, et al.
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
FILE REFERENCE: 802620-2005.1
CURRENT APPLICATION NUMBER: US/09/771,161A
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 09/724,676
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: 136776
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 135619
PRIOR FILING DATE: 2000-04-12
NUMBER OF SEQ ID NOS: 273
SOFTWARE: PatentIn version 3.0
SEQ ID NO: 173
LENGTH: 344
TYPE: PRT
ORGANISM: Homo sapiens
US-09-771-161A-173

Query Match 2.3%; Score 95.5; DB 10; Length 344;
Best Local Similarity 27.0%; Pred. No. 1.5;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDGPRVFLPIVEGFFRTSIOPGDDFVDVCV--ESSGSKVVDASFRLM 173
DB 184 ESSVSILQ---GARSPFCMPICE-----ISRGTINFSEELKIGSGGCGCYRAVMRNT 234
QY 174 LVL-----HAGDDPFALVEAMKIVRTHLGTFRLLLEKTPPGIVDKFGWCMTDAFYLVH 228
DB 235 YVAVKRLKENADLEWYAVKOSFLEVEQLSRFR-----HPNIVDPAGYCAONGFYCLVY 288
QY 229 ---PGVIEGVHRLVDGCGPR 246
DB 289 GFLPNGLSLDRHLHCOTQACPR 309

RESULT 10
US-09-773-753-5
Sequence 5, Application US/09773753
Patent No. US20020099177A1
GENERAL INFORMATION:
APPLICANT: NI, JIAN
FENG, PING
MUZIO, MARTA
DIXIT, VISHVA M.
NUMBER OF INVENTION: HUMAN IRAK-2
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn version #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/773,753
FILING DATE: 02-Feb-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/980,060
FILING DATE: <unknown>

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 712 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: No. US20020099177A1 Relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-773-753-5

Query Match
Best Local Similarity 27.0%; Score 95.5; DB 10; Length 712;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDSGRPYVFLPIVEGPFRTSIQPGDDFVDVCV--ESGSSKVVDAFSRSM 173
DB 184 ESSVSLQ---GARPSPFCPLCE-----ISRGTHNFSEELKIGEGFCGCVRAVWRNT 234
QY 174 LYL-----HAGDDPFALVKEAMKIVRTHLGTFRLLEKTPPGIVDKFGWCTWDAFYLTVA 228
DB 235 VYAVKRLKENADLEMTAVKOSFLEVEQLSRFR-----HPNIVDAGYCAQNGFCYCLVY 288
QY 229 ---POGVIEGVRHLVDGCGCP 246
DB 289 GFLPNSGLEDRHLHCOTQACPP 309

RESULT 11
US-09-771-161A-262
; Sequence 262, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 262
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-262

Query Match
Best Local Similarity 27.0%; Score 95.5; DB 10; Length 712;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDSGRPYVFLPIVEGPFRTSIQPGDDFVDVCV--ESGSSKVVDAFSRSM 173
DB 184 ESSVSLQ---GARPSPFCPLCE-----ISRGTHNFSEELKIGEGFCGCVRAVWRNT 234
QY 174 LYL-----HAGDDPFALVKEAMKIVRTHLGTFRLLEKTPPGIVDKFGWCTWDAFYLTVA 228
DB 235 VYAVKRLKENADLEMTAVKOSFLEVEQLSRFR-----HPNIVDAGYCAQNGFCYCLVY 288
QY 229 ---POGVIEGVRHLVDGCGCP 246
DB 289 GFLPNSGLEDRHLHCOTQACPP 309
```

```

;
; RESULT 12
; US-09-771-161A-263
; Sequence 263, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 263
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-263

Query Match
Best Local Similarity 27.0%; Score 95.5; DB 10; Length 712;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDSGRPYVFLPIVEGPFRTSIQPGDDFVDVCV--ESGSSKVVDAFSRSM 173
DB 184 ESSVSLQ---GARPSPFCPLCE-----ISRGTHNFSEELKIGEGFCGCVRAVWRNT 234
QY 174 LYL-----HAGDDPFALVKEAMKIVRTHLGTFRLLEKTPPGIVDKFGWCTWDAFYLTVA 228
DB 235 VYAVKRLKENADLEMTAVKOSFLEVEQLSRFR-----HPNIVDAGYCAQNGFCYCLVY 288
QY 229 ---POGVIEGVRHLVDGCGCP 246
DB 289 GFLPNSGLEDRHLHCOTQACPP 309

RESULT 13
US-09-771-161A-264
; Sequence 264, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 264
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-264

Query Match
Best Local Similarity 27.0%; Score 95.5; DB 10; Length 712;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDSGRPYVFLPIVEGPFRTSIQPGDDFVDVCV--ESGSSKVVDAFSRSM 173
```

```
Db 184 ESSVSLLQ---GARPSPFCWPLCE-----ISRGTHNFSSEELKIGGCGCYVAVAMRMT 234
Oy 174 LYL-----HAGDDPFLVKEAMKIYRTHLGFRLLEKTPPGIVDKFGMCTWDAFLYTH 228
Db 235 VYAVVRLEKEMADLEWYAVAQSFLEVEQLSRFR-----HPIYVDPAGYCAQNGFCILYV 288
Oy 229 ---POGVIEGVRHLVDGCGPP 246
Db 289 GFLPNGLSEDLRLHCQTQACPP 309

RESULT 14
US-10-101-464A-954
; Sequence 954, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020C2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; PRIOR FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 954
; LENGTH: 1021
; TYPE: PRT
; ORGANISM: EucaIyptus grandis
US-10-101-464A-954

Query Match 2.2%; Score 94; DB 9; Length 1021;
Best Local Similarity 20.7%; Pred. No. 11;
Matches 139; Conservative 97; Mismatches 217; Indels 220; Gaps 42;

Oy 164 KYVDSFRSMLYLHAGDDPFLVKEAMKIYRTHLGFRLLEKTPPGIVDKFGMCTWDAF 223
Db 152 RYLDLYNNM---TGDLPVA--VAEMPOLRHLHIG--NFGGIRP-----EYC--RMEI 199
Oy 224 -LYTH-----PO--GVIEGVRHL-----VDGCGPPGLVLIIDGWSIGHSDPI 265
Db 200 EYLAWSGDELGAIRPEIGNLFTLRILGYVNSYDGLRP-----EIGNLSQLY 249
Oy 266 TEGMNOYVAGGOMPRLKLFQENYKFRDYN-----FKATGPRAGOKGM-----KAFLD 315
Db 250 RUDAAAGSGSGE--IPEIKLLOMLDTLFLQVNGFAGSLPAEIGYLSLSDLSNMFGAG 308
Oy 316 ELKGEKTEYHVVYVWHALGCGWGLRPO--VPGDPEARVIO-----PYLSPLC-----OM 363
Db 309 EIPESLSQAKNLTLLHLFNNKLNESPERIADIRELOVQLWMNNTGSPGLGKNGHL 368
Oy 364 TMEDLAVDKIVLHKVGLVPRPEKAEMEYEGILAHLEKVGIDYKIDVILHLEMLCEDYGR 423
Db 369 QIVDLSNNKL---TGLTRP-----LC--YGNO 391
Oy 424 VDLAKYVYKMTKSIKHKRKNVIAEMHCNDFMLGTGATSLGVRGDDFWCTDPSGP- 482
Db 392 LQTLILSNVLLGPIRESL---GKCRSLEI-----RMGGNYL-----NGSI 430
Oy 483 PNCFTWLOGCHVHNCANSLMGNFTIHPDMFQSTHPCAAFAHAAASRAISGCPYVSD-- 540
Db 431 PRGLFELPELNOYE--FQDNILVGEFVPSD-----DSIALKLGSIITLSNKK 474
Oy 541 -----SVGKHNFLLKLVLPDGSILNSEYVALPTRDCLFEDPLHNGETMLKI--WNLN 592
```

```
Db 475 LCGSLPPTIG--NFSGVQKLLL--DGNFESGQ---IPPE-----IGRLQOLAKIDFSSN 521
Oy 593 KFTGVGARNCGGGGCRTRNOCFSQYSKRVSTNTNKDIEMH--SGENISIEGYVT 650
Db 522 RYSGPIPA-----QISQC-----KLLTF---VDLSGMLSGEINETIGMRI 560
Oy 651 FALYLYQAKKILISKPSODDIALDPFEFELITVSPVTKLIQTSLHFAPIGLVNMLNTSG 710
Db 561 LN-----YLNLSNNLSGSIRP-----SISTMOQL--TSVDFSYNNLSGLVPGTG 603
Oy 711 AIGSVYDDDLSSVEIGVCGEMRVFASKRPACRIDEDYGFYDQDMYVVOYVPMPI 770
Db 604 QFSYFNYTSFLGNPEL---CGPYL-----GPCK--DDEMGIHQ-----AHVKGPL 644
Oy 771 DSSSGISYIEVL 783
Db 645 SASVKLLVIGILL 657

RESULT 15
US-09-759-508B-2
; Sequence 2, Application US/09759508B
; Publication No. US2002018259A1
; GENERAL INFORMATION:
; APPLICANT: Fishman, Mark C.
; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
; FILE REFERENCE: 00786/381002
; CURRENT APPLICATION NUMBER: US/09/759,508B
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/175,787
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 26926
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-508B-2

Query Match 2.2%; Score 93; DB 9; Length 26926;
Best Local Similarity 19.9%; Pred. No. 2.3e+03;
Matches 97; Conservative 61; Mismatches 136; Indels 194; Gaps 28;

Oy 7 NGSNNVVF--GOLND-----MSSPFIIDGSDFTVNGHSFLSDVPEINVAS 50
Db 13413 DGGSRIIGYVDFLETEENKQRYMKSLSIQYSAKDLTEGKETIFR---VSAENENGEGT 13468
Oy 51 PSBYTSDIKSPVSVGCFVGFDASEPPDSRHVVSIGLKLDIRFM---SIFRFVY----- 99
Db 13469 PSREITVARD-----DVVAPD---LDLKGLPDLCYLAKENSFRLKPIKGRPA 13514
Oy 100 ---WTTTHWVRNGCD--LESETQIYI-----LEKSDSGRPYVFLPLPIVSEP 141
Db 13515 PSVSW-----KKGEDPLATDTRVSVESAVNTLLIYVDCQSDPAK--YTTTLKKNVAGT 13566
Oy 142 -----FRTSIOPGDDDFVDCVES-----GSSKYVDASFRLMLYHAGD 180
Db 13567 KEETSIKAVGKRGITGTGKIKDEYTAEMATLKMAPPKDDGSGSEITN-----YILEKR 13619
Oy 181 DPF-----ALVKEAMKIYRTHLG--TRRLLEKTPPGIVDKFGMCTWDAFLYTH 228
Db 13620 DSVNNKMWTCASAVQKTTFRVTRRLHEGMETPRVSAE-----N 13657
Oy 229 POCVITGV-----KH-----LVDGCGPPGLVLIIDGWSIGHSDPIETEGMNOYVAGEQ 278
Db 13658 KYGVGGLKSEPIVARHPEDVPDAPPPPIV-----DVRHDSVSLTWTDPKKT--GCGSP 13709
Oy 279 MPCRLLKFOEN-----YKFRDYNVPKATGPRRAGOKGMKAFIDELKGEFTVH 326
Db 13710 ITGYHLHEFERNSLMLKRRANKPIRMKDF---KYVTLTEG-----LEVEFVVM-- 13754
Oy 327 VYVWHALGCGWGLRPOVGLPEARVIOPLSPGLQMTMEDLAVDKIVLHKVGLVPRPEKA 386
```

Db 13755 -----AINLAGVGKPSLPSEPVPV-ALDPIDPPGKP-----EVINITRNSVTLLI---WT 13798
QY 387 EMEVEGLH 394
Db 13799 EPKIDGGH 13806

Search completed: April 4, 2003, 11:18:14
Job time : 47 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 14:59:10 ; Search time 7021 Seconds
(without alignments)
9749,295 Million cell updates/sec

Title: US-09-675-208-4_COPY_56_2407

Perfect score: 2352
Sequence: 1 ATGGCTCTAGTTTAAAAA.....CGTTATCGAGTACTGCTTT 2352

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: GenBank:*
- 2: gb_ha:*
- 3: gb_in:*
- 4: gb_cm:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rtd:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2352	100.0	2517	6 E15707	E15707 Cucumis sat
2	2352	100.0	2569	6 E25430	E25430 Rafilnose s
3	2352	100.0	2569	6 AF073744	AF073744 Cucumis s
4	1041.4	44.3	2690	6 E36417	E36417 Rafilnose s
5	976.4	41.5	2498	6 AR182382	AR182382 Sequence
6	975.4	41.5	2690	6 E371133	E371133 Rafilnose s
7	966.8	41.1	2343	6 BD010263	BD010263 Mutant pr
8	966	41.1	2497	6 E24424	E24424 Rafilnose s
9	956.6	40.7	2671	6 PSNA26475	PSNA26475 Pisum sat
10	956.6	40.7	2746	6 E24423	E24423 Rafilnose s
11	950.6	40.4	2395	6 AY081645	AY081645 Arabidops
12	950.6	40.4	2718	6 AY062781	AY062781 Arabidops
13	886.8	37.7	1762	6 E24425	E24425 Rafilnose s
14	866.4	36.8	1762	6 E36418	E36418 Rafilnose s
15	406.2	17.3	2727	6 PSNA311087	PSNA311087 Pisum sat
16	400.4	17.0	2592	6 SS1344091	SS1344091 Stachys s
17	387.6	16.5	2889	6 AME487030	AME487030 Alonsoa m
18	387	16.5	3073	6 VANI19024	VANI19024 Vigna angl
19	364	15.5	135295	8 AP003282	AP003282 Oryza sat
20	364	15.5	175174	8 AP003339	AP003339 Oryza sat
21	303.8	12.9	928	6 E30796	E30796 Rafilnose s
22	301.6	12.8	2714	6 AY090237	AY090237 Arabidops
23	275.8	11.7	2780	6 E25448	E25448 Rafilnose s
24	274.2	11.7	2738	6 AY050772	AY050772 Arabidops
25	263.8	11.2	86263	8 AB006702	AB006702 Arabidops
26	252.2	10.7	2669	8 BOBUTIN	BOBUTIN
27	237.6	10.1	2673	8 PM133148	PM133148 Petrea am
28	170.4	7.2	120185	8 AC007138	AC007138 Arabidops
29	170.4	7.2	198220	8 ATCHRIV5	ATCHRIV5
30	169.6	7.2	993	6 E24426	E24426 Rafilnose s
31	157.8	6.7	101458	8 F5024	F5024
32	130.8	5.6	1487	8 CAR271668	CAR271668 Arabidops
33	111.8	4.8	120868	2 AP003748	AP003748 Cicer ari
34	97	4.1	122095	8 AP002854	AP002854 Oryza sat
35	90.8	3.9	253	6 E25451	E25451 Rafilnose s
36	81.4	3.5	87503	6 ATT8H10	ATT8H10
37	78.4	3.3	850	8 AF159378	AF159378 Arabidops
38	71.8	3.1	4788	8 BLXSD1A	BLXSD1A Hordeum vul
39	71.2	3.0	125020	9 AF429315	AF429315 Homo sapi
40	70.6	3.0	109171	8 AC002328	AC002328 Genomic s
41	66	2.8	90767	2 H0510A06	AL442104 Oryza sat
42	64.4	2.7	129510	8 OSJN00047	AL606614 Oryza sat
43	64	2.7	154909	2 AP005174	AP005174 Oryza sat
44	64	2.7	162317	2 AP005148	AP005148 Oryza sat
45	56	2.4	1071	8 CASIMBIBP	X95875 C. arletium

ALIGNMENTS

RESULT 1
LOCUS E15707 2517 bp DNA linear PAT 28-JUL-1999
DEFINITION Cucumis sativus mRNA for raffinose synthase, complete cds.
ACCESSION E15707
VERSION E15707.1 GI:5710390
KEYWORDS JP 1998084973-A/1.
SOURCE JP 1998084973-A/1.
ORGANISM Cucumis sativus
REFERENCE 1 (bases 1 to 2517)
AUTHORS Oosumi C, Nozaki J, and Kida T.
TITLE RAFFINOSE SYNTHASE GENE, PRODUCTION OF RAFFINOSE AND TRANSFORMED
JOURNAL PATENT: JP 1998084973-A 1 07-APR-1998;
AJINOMOTO CO INC

COMMENT	OS	Cucumis sativus (cucumber)
	PN	JP 1998084973-A/1
	PD	07-APR-1998
	PR	28-APR-1997 JP 1997111124
	PF	26-APR-1996 JP 96P 107682, 26-JUL-1996 JP 96P 198079 PI
	PO	OSUMI CHERKO, NOZAKI JINJI, KIDA TAKAO
	PC	C12N15/09, A01H5/00, C12N9/00;
	CC	strandedness: Double;
	CC	topology: Linear;
	EH	Key
	EH	Location/Qualifiers
	FT	source
	FT	1..2517
	FT	/organism='Cucumis sativus'
	FT	/tissue_type='leaf'
	FT	56..2410
	CDS	/product='raffinose synthase'.
	FT	
FEATURES	source	Location/Qualifiers
		1..2517
		/organism="unidentified"
		/db_xref="taxon:32644"
BASE COUNT	613 a	524 c
ORIGIN		642 g 738 t
Query Match	100.0%;	Score 2352; DB 6; Length 2517;
Best Local Similarity	100.0%;	Pred. No. 0;
Matches 2352;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1	ATGGCTCTAGTTTAAATAATGCTGGCTCCAAACGATGTTTCAATTGATGGCTTAATAC 60
DB	56	ATGGCTCTAGTTTAAATAATGCTGGCTCCAAACGATGTTTCAATTGATGGCTTAATAC 115
OY	61	ATGTGTCACGGTTTGCATCGACGGATCGATTTCACTGTGAACGGTCATTCTTCTG 120
DB	116	ATGTGTCACGGTTTGCATCGACGGATCGATTTCACTGTGAACGGTCATTCTTCTG 175
OY	121	TCCGATGTTCCGTGAGAAATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
DB	176	TCCGATGTTCCGTGAGAAATGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
OY	181	CCGGTTTCGGTTGGTGGCTTGTGATTCGACGGCTGGGAACCTGATAGCCGACATGTT 240
DB	236	CCGGTTTCGGTTGGTGGCTTGTGATTCGACGGCTGGGAACCTGATAGCCGACATGTT 295
OY	241	GTTTGGATTTGGGAGCTGAAGGATATTTGGCTTTATGAGTATTTTCAGTTTAAGCTTTGG 300
DB	296	GTTTGGATTTGGGAGCTGAAGGATATTTGGCTTTATGAGTATTTTCAGTTTAAGCTTTGG 355
OY	301	TGGACTACACACTGGGTTGTCGAAATGCTGGGGATCTTGATCGAGACTCAGATTGTG 360
DB	356	TGGACTACACACTGGGTTGTCGAAATGCTGGGGATCTTGATCGAGACTCAGATTGTG 415
OY	361	ATCCTTGAAGATCAGATTGCTGTCGACCGATGTTTCTTCTTCGATCGTTGAGGGA 420
DB	416	ATCCTTGAAGATCAGATTGCTGTCGACCGATGTTTCTTCTTCGATCGTTGAGGGA 475
OY	421	CCGTTCCGAACCTGATTCAGCTCGCTGGGGATGATGACTTTGCGATGGTTGTGTGAGAGT 480
DB	476	CCGTTCCGAACCTGATTCAGCTCGCTGGGGATGATGACTTTGCGATGGTTGTGTGAGAGT 535
OY	481	GCTTGTGGAAGTTGTGTGATGATCGTTCGAAATATGTTGATATCTTCAGCTGCTGTAT 540
DB	536	GCTTGTGGAAGTTGTGTGATGATCGTTCGAAATATGTTGATATCTTCAGCTGCTGTAT 595
OY	541	GATCCGTTTGCATCTTTAAAGAGGCGATGAAGATCGTGAGAGCCCATCTTGAACCTTTT 600
DB	596	GATCCGTTTGCATCTTTAAAGAGGCGATGAAGATCGTGAGAGCCCATCTTGAACCTTTT 655
OY	601	CGCTGTGTGAGAGGAGACTCCACAGATATCGTGAGCAATTCGGTTGGTGTGACGGTG 660
DB	656	CGCTGTGTGAGAGGAGACTCCACAGATATCGTGAGCAATTCGGTTGGTGTGACGGTG 715
OY	661	GACGGTTTACCTAAGCGTTTCATCCACAGGCGTAAATAGAAGCGTGAGGCATCTCTGTC 720

Dh	716	GACGCGTTTAACTAACGGTTTCATCCACAGGGCGTAAATGAAAGCGTGAAGCATCTCGTC	773
Qy	721	GACGGCGGTTCTCTCCCGGGTTTAGTCTTAATGACAGATGGTGGCAATTCATTCGAGAC	780
Dh	776	GACGGCGGTTCTCTCCCGGGTTTAGTCTTAATGACAGATGGTGGCAATTCATTCGAGAC	835
Qy	781	GATTTCGATTCCTCATCCCAAAAGAAATACCAACCGTTCGGCGGCGACGAATTCGCC	840
Dh	836	GATTTCGATTCCTCATCCCAAAAGAAATACCAACCGTTCGGCGGCGACGAATTCGCC	895
Qy	841	TGCGCTTTTGGAAATTCCAAGAAATATACAAATTCGGTACATCTGCATCCCAAGGCC	900
Dh	896	TGCGCTTTTGGAAATTCCAAGAAATATACAAATTCGGTACATCTGCATCCCAAGGCC	955
Qy	901	ACCGGCCCCCGAGCGCCGACAGAGGGATGAGCGCTTATATGATGACTCAAGAGAGC	960
Dh	956	ACCGGCCCCCGAGCGCCGACAGAGGGATGAGCGCTTATATGATGACTCAAGAGAGC	1015
Qy	961	TTTAAGACTGTGGAGCAATGTTATGTTTGGCATGCTTGTGGTGAATTTGGGTGGCCCT	1020
Dh	1016	TTTAAGACTGTGGAGCAATGTTATGTTTGGCATGCTTGTGTGGATTTGGGTGGCCCT	1075
Qy	1021	CGCGCGAGGTGCCTGGCTTGCCCTGAGGCAAGGTGATTTAGCCAGTGCCTTTCACAGG	1080
Dh	1076	CGCGCGAGGTGCCTGGCTTGCCCTGAGGCAAGGTGATTTAGCCAGTGCCTTTCACAGG	1135
Qy	1081	CTGAGATGACAGATGGAGAAATTTGGCGCGTGGATACAAATGTTCTTCATTAAGTTCGGCGC	1140
Dh	1136	CTGAGATGACAGATGGAGAAATTTGGCGCGTGGATACAAATGTTCTTCATTAAGTTCGGCGC	1195
Qy	1141	GTCGCGCGAGAAAGCGCTGAGGAGATGTACGAAGCACTTCATGCTCATTTTGGAAAAAGTT	1200
Dh	1196	GTCGCGCGAGAAAGCGCTGAGGAGATGTACGAAGCACTTCATGCTCATTTTGGAAAAAGTT	1255
Qy	1201	GGGATCCAGCGGTGTAAAGATTGACAGTTACACCTTGTGGAGATGTTGTGTCGAACATAT	1260
Dh	1256	GGGATCCAGCGGTGTAAAGATTGACAGTTACACCTTGTGGAGATGTTGTGTCGAACATAT	1315
Qy	1261	GGAGGAGAGATGGATTTTGGCAAAAGGCAATATACAAAGCATGACCAAAATCAATTAATATA	1320
Dh	1316	GGAGGAGAGATGGATTTTGGCAAAAGGCAATATACAAAGCATGACCAAAATCAATTAATATA	1375
Qy	1321	CATTTTAAAGAAATGAGATCATTTGCAAGTATGAAACATTGTAGACATCTCATGTTCTTT	1380
Dh	1376	CATTTTAAAGAAATGAGATCATTTGCAAGTATGAAACATTGTAGACATCTCATGTTCTTT	1435
Qy	1381	GGCAGCGAGCATCTCTCTTGGCGGTTGGTGGATGACATTTTGGCGACGAGACCCCTCT	1440
Dh	1436	GGCAGCGAGCATCTCTCTTGGCGGTTGGTGGATGACATTTTGGCGACGAGACCCCTCT	1495
Qy	1441	GGTGATCCAAACGGTATCGTTTGGCTCCAAAGATGTACATGGTTCATTTGTCCAAAGAC	1500
Dh	1496	GGTGATCCAAACGGTATCGTTTGGCTCCAAAGATGTACATGGTTCATTTGTCCAAAGAC	1555
Qy	1501	AGCTTGGAGTGGGGAGCTTCATCCACCCTGATGGGATATGTGCCAATTCGACCCACCT	1560
Dh	1556	AGCTTGGAGTGGGGAGCTTCATCCACCCTGATGGGATATGTGCCAATTCGACCCACCT	1615
Qy	1561	TGTGCGCGCTTCATGCTGCGCTCTCGAGCCATCTGTTGGCGCGCATCTATGTTAGTAT	1620
Dh	1616	TGTGCGCGCTTCATGCTGCGCTCTCGAGCCATCTGTTGGCGCGCATCTATGTTAGTAT	1675
Qy	1621	TCTGTGGGAAAGCTATACCTTTGATCTTCTTGAAAAACCTAGTCTCTTCGATGAGATCGATC	1680
Dh	1676	TCTGTGGGAAAGCTATACCTTTGATCTTCTTGAAAAACCTAGTCTCTTCGATGAGATCGATC	1735
Qy	1681	CTTGGAGTGAATCTATCGATCCGACCTCGAGATTTGTTTGTGAAGACCTTTTGAT	1740
Dh	1736	CTTGGAGTGAATCTATCGATCCGACCTCGAGATTTGTTTGTGAAGACCTTTTGAT	1795
Qy	1741	AATGGAGAACTATGCTTAAAGATTTGGAAATCTCACACAGTTCAGTGAAGTATGTTGTCCA	1800
Dh	1796	AATGGAGAACTATGCTTAAAGATTTGGAAATCTCACACAGTTCAGTGAAGTATGTTGTCCA	1855

Accession	Source	Organism	Reference Authors Title	Journal	Comment
QY 1801	TTCAACTGCGCAAGAGAGAGATGCTGTGCTGAGAGACACGCCCAACCAATGCTTTTCACAA	1860			
Db 1856	TTCAACTGCCCAAGGAGGAGATGCTGTGCTGAGACACGCCCAACCAATGCTTTTCACAA	1915			
QY 1861	TACTCMAAAGCAGTGCATCTCCAAAACCTAACCCMAAAGCATTGAAATGGCCACAGTGGACAA	1920			
Db 1916	TACTCMAAAGCAGTGCATCTCCAAAACCTAACCCMAAAGCATTGAAATGGCCACAGTGGACAA	1975			
QY 1921	AACCTATCTCTAATTGAAGGGCTTTAAACCTTTGGGCTTTACCTCTATCAAGCCAAAAA	1980			
Db 1976	AACCTATCTCTAATTGAAGGGCTTTAAACCTTTGGGCTTTACCTCTATCAAGCCAAAAA	2035			
QY 1981	CTTATCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTGGAAATTGCAG	2040			
Db 2036	CTTATCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTGGAAATTGCAG	2095			
QY 2041	CTCATCACTGTTTTCACGAGTGCACCAACTCATCCAACTTCTCTACACTTTGGCCCAATT	2100			
Db 2096	CTCATCACTGTTTTCACGAGTGCACCAACTCATCCAACTTCTCTACACTTTGGCCCAATT	2155			
QY 2101	GGGCGTGGTGAACATGCTTTAAGACTAGTGGAGCCATCATCTGTGGACTATGACGATGAC	2160			
Db 2156	GGGCGTGGTGAACATGCTTTAAGACTAGTGGAGCCATCATCTGTGGACTATGACGATGAC	2215			
QY 2161	CTAAGCTCAGTCGACATTTGGTGTCAAAGGGTGTGTGAGATGCGAGATTTTGCATCGAA	2220			
Db 2216	CTAAGCTCAGTCGACATTTGGTGTCAAAGGGTGTGTGAGATGCGAGATTTTGCATCGAA	2275			
QY 2221	AAACCAAGGGCTTGCTGATTTGATGATGGGAGAGATTTGGGTTCAACTATGATCAGGACCA	2280			
Db 2276	AAACCAAGGGCTTGCTGATTTGATGATGGGAGAGATTTGGGTTCAACTATGATCAGGACCA	2335			
QY 2281	ATGCTGTGTGCTTCAAGTGCATGCGCCAAATTGATCTTCATCGGGTGGCATTTGCGTTATC	2340			
Db 2336	ATGCTGTGTGCTTCAAGTGCATGCGCCAAATTGATCTTCATCGGGTGGCATTTGCGTTATC	2395			
QY 2341	GAGTACTGTTT 2352				
Db 2396	GAGTACTGTTT 2407				
RESULT 2					
E25430					
LOCUS	E25430	2569 bp	DNA	linear	PAT 18-JUN-2001
DEFINITION	Raflinose synthase gene, process for producing raflinose and				
	transformed plant.				
ACCESSION	E25430				
VERSION	E25430.1	GI:13024726			
KEYWORDS	JP 1999123080-A/1.				
SOURCE	Cucumis sativus.				
ORGANISM	Cucumis sativus.				
	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.				
REFERENCE	Chieko, O., Jinji, N. and Takao, K.				
AUTHORS	Raflinose synthase gene, process for producing raflinose and				
TITLE	transformed plant				
JOURNAL	Patent: JP 1999123080-A 1 11-MAY-1999;				
COMMENT	AJINOMOTO CO INC				
	OS Cucumis sativus (Cucumber)				
	PN JP 1999123080-A/1				
	PD 11-MAY-1999				
	PE 24-OCT-1997 JP 1997292969				
	PR				
	PI CHIEKO OZUMI, JINJI NOZAKI, TAKAO KIDA				
	PC C12N15/09, A01HS/00, C12N9/00, C12N1/19, C12N15/00 CC				
	Strandedness: Double;				
	CC Topology: Linear;				
	EH Key				
FEATURES	FT CDS				
	Location/Qualifiers				
	56..2407.				

Source	1. 2569	Organism="Cucumis sativus"
BASE COUNT	641 a 527 c 643 g 755 t	3 others
ORIGIN		
Query Match	100.0%;	Score 2352; DB 6; Length 2569;
Best Local Similarity	100.0%;	Pred. No. 0;
Matches 2352; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

QY	1	ATNGGCTCCTAGTGTAAAAATGTCGGCTCCACACGATGTTCAATTTGATGCGCTTAATGAC	60
DB	56	ATGGCTCCCTAGTGTAAAAATGTCGGCTCCACACGATGTTCAATTTGATGCGCTTAATGAC	115
QY	61	ATGTCGTCACGTTTGGCAATCGACGATCGATTCACGTCGTAACGGTCATTCGTTTTCG	120
DB	116	ATGTCGTCACGTTTGGCAATCGACGATCGATTCACGTCGTAACGGTCATTCGTTTTCG	175
QY	121	TCCGATGTTCTGAGAACATTTGTTGCTTCTCCTTCCGTACACTTGCATAGACAATCC	180
DB	176	TCCGATGTTCTGAGAACATTTGTTGCTTCTCCTTCCGTACACTTGCATAGACAATCC	235
QY	181	CCGGTTTGGGTTGGTTGCTTGTGTTGGATTGCACGGGTGGAAACCTGATAGCCGACATGTT	240
DB	236	CCGGTTTGGGTTGGTTGCTTGTGTTGGATTGCACGGGTGGAAACCTGATAGCCGACATGTT	295
QY	241	GTTTGGATTGGGAGACTGAAGATATTGCGGTTATGAGATTTTCAGGTTTAAAGGTTTGG	300
DB	296	GTTTGGATTGGGAGACTGAAGATATTGCGGTTATGAGATTTTCAGGTTTAAAGGTTTGG	355
QY	301	TGGACTACACACTGGGTTGGTTCGAAATGGTGGGATCTTGAATCGAGACTGAGATTGTG	360
DB	356	TGGACTACACACTGGGTTGGTTCGAAATGGTGGGATCTTGAATCGAGACTGAGATTGTG	415
QY	361	ATTCCTTGAGAACTGACGATTCCTGTGACCGCTATGTTTCTCTTCCCATCGTTGAGGA	420
DB	416	ATTCCTTGAGAACTGACGATTCCTGTGACCGCTATGTTTCTCTTCCCATCGTTGAGGA	475
QY	421	CCGTTCCGAAACCTCGATTGAGCTGGGGATGATGACTTTGTCGATGTTTGTGTCGAGAT	480
DB	476	CCGTTCCGAAACCTCGATTGAGCTGGGGATGATGACTTTGTCGATGTTTGTGTCGAGAT	535
QY	481	GGTTGGTCGAAAGTTGTTGATGATCGTTCGAAAGTATGTTGTATCTTCATGCTGGTAT	540
DB	536	GGTTGGTCGAAAGTTGTTGATGATCGTTCGAAAGTATGTTGTATCTTCATGCTGGTAT	595
QY	541	GATTCGTTTGCACCTGTTTAAAGAGCGGATGAAGATCTGTAGAGACCCATCTTGGAACTTTT	600
DB	596	GATTCGTTTGCACCTGTTTAAAGAGCGGATGAAGATCTGTAGAGACCCATCTTGGAACTTTT	655
QY	601	CGCTTGTGAGAGAGAACCTCACAGGATTCGTGAGAACAAATTCGGTTGGTGCACGTGG	660
DB	656	CGCTTGTGAGAGAGAACCTCACAGGATTCGTGAGAACAAATTCGGTTGGTGCACGTGG	715
QY	661	GACGCGTTTTACCTAACGCTTTCATCCACAGGGCGTAAATGAAGAGCGTGGAGATCTGTC	720
DB	716	GACGCGTTTTACCTAACGCTTTCATCCACAGGGCGTAAATGAAGAGCGTGGAGATCTGTC	775
QY	721	GACGCGGTTTGTCTCCCGGTTTATGCTTAATCGACGATGTTGGCAATTCATCGGACAC	780
DB	776	GACGCGGTTTGTCTCCCGGTTTATGCTTAATCGACGATGTTGGCAATTCATCGGACAC	835
QY	781	GATTCGATTCCTCATACCAAGAAGAAATGAAGCAAAACCGTGGCGGCGACGAATGCC	840
DB	836	GATTCGATTCCTCATACCAAGAAGAAATGAAGCAAAACCGTGGCGGCGACGAATGCC	895
QY	841	TCCCGCTTTTGAATTTCCAAAGAAATTTCAAAATTCCTGTACGTACGTCAATCCCAAGGCC	900
DB	896	TCCCGCTTTTGAATTTCCAAAGAAATTTCAAAATTCCTGTACGTACGTCAATCCCAAGGCC	955
QY	901	ACCGCGCCCCGAGCGCGCCAGAGGGGATGAAGGCGCTTTATGATGTAACCTCAAGAGAG	960
DB	956	ACCGCGCCCCGAGCGCGCCAGAGGGGATGAAGGCGCTTTATGATGTAACCTCAAGAGAG	1015

QY 961 TTTAAGACTGTGGAGCATGTTTAATGTTTGGCAGTCCTTGTGTGATATTTGGGGTGGCCTT 1020
 |||||
 Db 1016 TTTAAGACTGTGGAGCATGTTTAATGTTTGGCAGTCCTTGTGTGATATTTGGGGTGGCCTT 1075
 |||||
 QY 1021 CGCCCGAGTGGCTTGGCTTGGCTGAGGACGTGTGATTCAGCCAGTGCCTTTCACACGAGG 1080
 |||||
 Db 1076 CGCCCGAGTGGCTTGGCTTGGCTGAGGACGTGTGATTCAGCCAGTGCCTTTCACACGAGG 1135
 |||||
 QY 1081 CTGAGATGAGTGAAGGAGTGGTGGGATTAAGATTTCTTCTCAATAGTCGCGCTG 1140
 |||||
 Db 1136 CTGAGATGAGTGAAGGAGTGGTGGGATTAAGATTTCTTCTCAATAGTGGGCTG 1195
 |||||
 QY 1141 GTCCCGCGGAGAGGCTGAGAGATGTACGAGGACTTCATGCTCATTTGGAAAAAGTT 1200
 |||||
 Db 1196 GTCCCGCGGAGAGGCTGAGAGATGTACGAGGACTTCATGCTCATTTGGAAAAAGTT 1255
 |||||
 QY 1201 GGGATCGACGCTGTAAAGATTGACGTTATCCACCTATTTGAGATGTTGTGAAAGACTAT 1260
 |||||
 Db 1256 GGGATCGACGCTGTAAAGATTGACGTTATCCACCTATTTGAGATGTTGTGAAAGACTAT 1315
 |||||
 QY 1261 GAGAGGAGATGGATTGGCAAGGCATTTACAAAGCATGACCAATCAATTAATATAA 1320
 |||||
 Db 1316 GAGAGGAGATGGATTGGCAAGGCATTTACAAAGCATGACCAATCAATTAATATAA 1375
 |||||
 QY 1321 CATTTTAAAGAAATGAGTCAATTTGCAAGTATGGAACATTTGAACGACTTCATGTTCTT 1380
 |||||
 Db 1376 CATTTTAAAGAAATGAGTCAATTTGCAAGTATGGAACATTTGAACGACTTCATGTTCTT 1435
 |||||
 QY 1381 GGCACGGAAGCTATCTCTCTTGGTCCGTTGGTGGATGATGATCTTTGGTGACAGCACCTCT 1440
 |||||
 Db 1436 GGCACGGAAGCTATCTCTCTTGGTCCGTTGGTGGATGATGATCTTTGGTGACAGCACCTCT 1495
 |||||
 QY 1441 GGTGATCCAAACGAGTGTGGTGGTCCAAAGATGTCACATGATGTTGTCATTTGGCAACGAC 1500
 |||||
 Db 1496 GGTGATCCAAACGAGTGTGGTGGTCCAAAGATGTCACATGATGTTGTCATTTGGCAACGAC 1555
 |||||
 QY 1501 AGCTTGTGATGGGGAACCTTCATCCACCCTGACTGGGATATGTTCCAAATCCACCACCCT 1560
 |||||
 Db 1556 AGCTTGTGATGGGGAACCTTCATCCACCCTGACTGGGATATGTTCCAAATCCACCACCCT 1615
 |||||
 QY 1561 TGTGCGCCCTTCATGCTGCTCTGAGGCCATCTGTGTGGCCCGAGTCTTGTAGTATGAT 1620
 |||||
 Db 1616 TGTGCGCCCTTCATGCTGCTCTGAGGCCATCTGTGTGGCCCGAGTCTTGTAGTATGAT 1675
 |||||
 QY 1621 TCTGTGGGAAGCATTAATTTGATCTTCTGAAAAAACTAGTGCCTCTCTGATGATGATC 1680
 |||||
 Db 1676 TCTGTGGGAAGCATTAATTTGATCTTCTGAAAAAACTAGTGCCTCTCTGATGATGATC 1735
 |||||
 QY 1681 CTTCGAAAGTATGATATGACATCCGACCTCGGATTTGTTGTTGAAGACCCTTTGAT 1740
 |||||
 Db 1736 CTTCGAAAGTATGATATGACATCCGACCTCGGATTTGTTGTTGAAGACCCTTTGAT 1795
 |||||
 QY 1741 AATGGAAGAACTATGCTTAAGATTTTGAATCTCAACAAGTTCATGGAGTATTTGGTCA 1800
 |||||
 Db 1796 AATGGAAGAACTATGCTTAAGATTTTGAATCTCAACAAGTTCATGGAGTATTTGGTCA 1855
 |||||
 QY 1801 TTCACTGCCAAGAGAGAGATGTTCTGTGAGACAGCCGCAACCAATGCTTTTGACAA 1860
 |||||
 Db 1856 TTCACTGCCAAGAGAGAGATGTTCTGTGAGACAGCCGCAACCAATGCTTTTGACAA 1915
 |||||
 QY 1861 TACTCAAAAAGATGATTCAAAAACTAACCCAAAAAGACATAGAAATGGCAGATGGAGAA 1920
 |||||
 Db 1916 TACTCAAAAAGATGATTCAAAAACTAACCCAAAAAGACATAGAAATGGCAGATGGAGAA 1975
 |||||
 QY 1921 AACCCATATCTATTAAGGCGTTAAACCTTGGCTTTACTCTATCAAGCCAAAAA 1980
 |||||
 Db 1976 AACCCATATCTATTAAGGCGTTAAACCTTGGCTTTACTCTATCAAGCCAAAAA 2035
 |||||
 QY 1981 CTATATCTCTCCAAAGCCTCTCAAGATCTTGAATAGCTCTTGACCATTCGAATTCGAG 2040
 |||||
 Db 2036 CTATATCTCTCCAAAGCCTCTCAAGATCTTGAATAGCTCTTGACCATTCGAATTCGAG 2095
 |||||

QY 2041 CTCATCACTGTTTACCAGTGAACCAATCATCAACTTCTACACTTTGCCCAATT 2100
 |||||
 Db 2096 CTCATCACTGTTTACCAGTGAACCAATCATCAACTTCTACACTTTGCCCAATT 2155
 |||||
 QY 2101 GGGCTGTTGAACATGCTTAACTAGTGAAGCCATCCATCTGTGACATGACATGAC 2160
 |||||
 Db 2156 GGGCTGTTGAACATGCTTAACTAGTGAAGCCATCCATCTGTGACATGACATGAC 2215
 |||||
 QY 2161 CTAACTCACTGAGATTTGGTGTCAAAAGGTTGGTGTGAGATTTGGATTCGAA 2220
 |||||
 Db 2216 CTAACTCACTGAGATTTGGTGTCAAAAGGTTGGTGTGAGATTTGGATTCGAA 2275
 |||||
 QY 2221 AAACCAAGGCTTGTCTATTTGATGGGAGAGATTTGGTGTCAAGTATGATCAGACAA 2280
 |||||
 Db 2276 AAACCAAGGCTTGTCTATTTGATGGGAGAGATTTGGTGTCAAGTATGATCAGACAA 2335
 |||||
 QY 2281 ATGGTGTGTTCAAGTGGCAGCCATTTGATTTTCATGGGTGATTTGGTTATC 2340
 |||||
 Db 2336 ATGGTGTGTTCAAGTGGCAGCCATTTGATTTTCATGGGTGATTTGGTTATC 2395
 |||||
 QY 2341 GAGTACTTGT 2352
 |||||
 Db 2396 GAGTACTTGT 2407
 |||||

RESULT 3
 AF073744 2569 bp mRNA linear PLN 05-JAN-1999
 LOCUS Cucumis sativus raffinose synthase (Rfs) mRNA, complete cds.
 DEFINITION AF073744
 ACCESSION AF073744
 VERSION AF073744.1 GI:4106394
 KEYWORDS
 SOURCE Cucumis sativus.
 ORGANISM Cucumis sativus.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
 Rosidae; eustosids I; Cucurbitales; Cucurbitaceae; Cucumis.
 1 (bases 1 to 2569)
 AUTHORS Ohsumi, C., Nozaki, J.
 TITLES Direct Submission
 JOURNAL Submitted (22-JUN-1998) Central Research Laboratories, Ajinomoto
 Co., Inc., 1-1, Suzuki-cho, Kawasaki-ku, Kawasaki 210-8681, Japan

FEATURES
 source location/Qualifiers
 1..2569
 /organism="Cucumis sativus"
 /db_xref="taxon:3659"
 /tissue_type="leaf"
 /dev_stage="mature"
 1..2569
 /gene="Rfs"
 56..2410
 /gene="Rfs"
 /EC_number="2.4.1.82"
 /codon_start=1
 /product="raffinose synthase"
 /protein_id="AAD02832.1"
 /translation="MAPSEKNGSNVVSFDGLNDMSFPAIDSGFTVNGSHSLSDVP
 ENIVASPSPTYSIDKSPVSGVGEFDESEPSRHSVLSGLKDIEMSIPEKXVMT
 THWVRNGDLESFQIVILKSDSGRPVLLPIVEGPRFHSIQPCDDPVDVGVES
 GSKKYVDASFRLMILYHAGDDPFLAYKEMKIVRHLGTFRLLEKTPPGIYDKGWC
 TWDAVLYVHFGVIEGVRHVLVDSGCPPELVLDGWSISIGHSDPITKEGNQTVAG
 EQMPCRLKFDENYKFRDYVNPKATGPAGOGKMAFIDELKGEFTVEHYVYHVALC
 GYWGSLRPQVLPRLPARYIOPVLSGLQTMEDLAVDKIVLHKVGLVPEKAAEEMV
 LHAHLKVGIDGVKIDVHILHMLCEDYGGARDLAKAYKAMTKSINKHFGNGVYIAS
 MEHCNDPMLTGLTEALISLRVGDPECTPDSGDPNCTPMIOGHVHCANDSIAMNCTI
 HPDWMFQSTHPCAFHAASRAISGCPITYSDSVGKHNHFDLTKLVLPGSLIRSEYV
 ALPTDCLFEEDPLHNGETMLKIMNLNKTGVYGAFCNCGGCGRETRNOCFSQYSKR
 VTSKTPKDIEMHSGENPISIEGVTFALYLYQAKLILSKPSODLIDLPFEFELI
 TVSPVTKLIQTSILHFAPIGLVNLMTSGAISOVDVDDLSVETIGVKCGEMRVPASK
 KPRACRIGEDENVKPYDDQDVYVNPMPIDISSSGISIVETLP"

BASE COUNT 641 a 527 c 643 g 755 t 3 others
 ORIGIN

Query Match 100.0%: Score 2352; DB 8; Length 2569;
Best Local Similarity 100.0%: Pred. No. 0;
Matches 2352: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGCTCTAGTTTAAAAATGGTGGCTCCCAAGTATTTCATTGATGGCTTAATATGAC 60
DB 56 ATGCGCTCTAGTTTAAAAATGGTGGCTCCCAAGTATTTCATTGATGGCTTAATATGAC 115
QY 61 ATGTGCTCAACGCTTTGGCAATGACAGGATCGGATTTCACTGTGAACGGTCACTTTCTG 120
DB 116 ATGTGCTCAACGCTTTGGCAATGACAGGATCGGATTTCACTGTGAACGGTCACTTTCTG 175
QY 121 TCCGATGTTCCGAGAACATTTGCTCTCTCTCCGTACACTTTCGATATAGCAAGTCC 180
DB 176 TCCGATGTTCCGAGAACATTTGCTCTCTCTCTCCGTACACTTTCGATATAGCAAGTCC 235
QY 181 CCGGTTTGGGTTGTTGCTTTGTTGGATTTCGACGGCTGGGAACCTGATAGCCGACATGT 240
DB 236 CCGGTTTGGGTTGTTGCTTTGTTGGATTTCGACGGCTGGGAACCTGATAGCCGACATGT 295
QY 241 GTTTGCAATTTGGGAACCTGAGGATATTGGGTTTATGAGATTTTTCAGTTTAAGTTTGG 300
DB 296 GTTTGCAATTTGGGAACCTGAGGATATTGGGTTTATGAGATTTTTCAGTTTAAGTTTGG 355
QY 301 TGGACTACACACTGGGTTGGTGAATGTTGGGATCTTGAATCGGAGACTGAGATTGTG 360
DB 356 TGGACTACACACTGGGTTGGTGAATGTTGGGATCTTGAATCGGAGACTGAGATTGTG 415
QY 361 ATCTTGAAGAACAGATTCTGGTGCAGCCGTATGTTTTCCTTCTTCGATCGTTGAGGA 420
DB 416 ATCTTGAAGAACAGATTCTGGTGCAGCCGTATGTTTTCCTTCTTCGATCGTTGAGGA 475
QY 421 CCGTTCCTCAACCTCGATTCAGCCTGGGATGATGACTTTGTGATGTTGTGTCCAGAGT 480
DB 476 CCGTTCCTCAACCTCGATTCAGCCTGGGATGATGACTTTGTGATGTTGTGTCCAGAGT 535
QY 481 GGTTCGTGGAAGTTGTTGATGATCATGTTCCGAAGTATGTTGATCTTCATCTGATGAT 540
DB 536 GGTTCGTGGAAGTTGTTGATGATCATGTTCCGAAGTATGTTGATCTTCATCTGATGAT 595
QY 541 GATCCGTTTGGCACTTTGTTAAGAGCGGATGAAGATCGTGAGAGCCCATCTTGGAACTTTT 600
DB 596 GATCCGTTTGGCACTTTGTTAAGAGCGGATGAAGATCGTGAGAGCCCATCTTGGAACTTTT 655
QY 601 GCCTTGTGGAGAGAAGCTCCACAGGTATCGTGGCAAAATTTGGTTGGTGCACGTGG 660
DB 656 GCCTTGTGGAGAGAAGCTCCACAGGTATCGTGGCAAAATTTGGTTGGTGCACGTGG 715
QY 661 GACGCGTTTACCTAAGCGTTTCATCCACAGGGCGTAAATGAGAGCGTGAGCATGTCGT 720
DB 716 GACGCGTTTACCTAAGCGTTTCATCCACAGGGCGTAAATGAGAGCGTGAGCATGTCGT 775
QY 721 GACGCGGTTTGTCTCCGCGTTTGTAGTCTTAATGACGATGTTGGCAATTCATCGGACAC 780
DB 776 GACGCGGTTTGTCTCCGCGTTTGTAGTCTTAATGACGATGTTGGCAATTCATCGGACAC 835
QY 781 GATTGGGATCCCATCAACGAAGAAGTAAGCAAAACCGTGGCGGAGAGAATGGCC 840
DB 836 GATTGGGATCCCATCAACGAAGAAGTAAGCAAAACCGTGGCGGAGAGAATGGCC 895
QY 841 TGCCTCTTTTGAATTTCCAAAGATAATTAACAATTCGCTGACTACGTCAATCCCAAGGC 900
DB 896 TGCCTCTTTTGAATTTCCAAAGATAATTAACAATTCGCTGACTACGTCAATCCCAAGGC 955
QY 901 ACCGCGCCCCGAGCCGCGCAGAGGGAGTGAAGCGCTTTATAGATGAACCTCAAGAGAG 960
DB 956 ACCGCGCCCCGAGCCGCGCAGAGGGAGTGAAGCGCTTTATAGATGAACCTCAAGAGAG 1015
QY 961 TTTAAGACGTGAGAGCATGTTATGTTGGCATGCTTTGTGTGGATATTTGGGGTGGCCTT 1020
DB 1016 TTTAAGACGTGAGAGCATGTTATGTTGGCATGCTTTGTGTGGATATTTGGGGTGGCCTT 1075

QY 1021 CCCCCAGAGTCCCTGCTGGCTGAGCAGCATGTGATTCAGCCAGTCTTTACACAGG 1080
DB 1076 CCCCCAGAGTCCCTGCTGGCTGAGCAGCATGTGATTCAGCCAGTCTTTACACAGG 1135
QY 1081 CTGCAGATGACGATGGAGGATTTGGCGGTGATAGATTTGTTCTTCAATAGCTGGGCTG 1140
DB 1136 CTGCAGATGACGATGGAGGATTTGGCGGTGATAGATTTGTTCTTCAATAGCTGGGCTG 1195
QY 1141 GTCCCGCCGGAGAAAGCTGAGAGATGTACGAAGACCTTCATGCTCATTTTGGAAAAATT 1200
DB 1196 GTCCCGCCGGAGAAAGCTGAGAGATGTACGAAGACCTTCATGCTCATTTTGGAAAAATT 1255
QY 1201 GGGATCGACGCTTAAATGATGACCTTATCCACATTTGGAGATGTTGTGTAAGACTAT 1260
DB 1256 GGGATCGACGCTTAAATGATGACCTTATCCACATTTGGAGATGTTGTGTAAGACTAT 1315
QY 1261 GGAGGAGATGAGATTTGGCAAAAGCATATTACAAAGCAATGACCAAAATCAATAAATAA 1320
DB 1316 GGAGGAGATGAGATTTGGCAAAAGCATATTACAAAGCAATGACCAAAATCAATAAATAA 1375
QY 1321 CATTTTAAAGAAATGAGACTATTCGACATATGGAACATTTGTAAGACTTCACTGCTT 1380
DB 1376 CATTTTAAAGAAATGAGACTATTCGACATATGGAACATTTGTAAGACTTCACTGCTT 1435
QY 1381 GGCACGAAAGCTATCTCTTGGTGTGTGATGACTTTTGGTGGACGAGCCCTCT 1440
DB 1436 GGCACGAAAGCTATCTCTTGGTGTGTGATGACTTTTGGTGGACGAGCCCTCT 1495
QY 1441 GGTATTCGAACGCTATCTTTGGCTCCAAAGATGTCAATGTTCAATTTGCAAGCAG 1500
DB 1496 GGTATTCGAACGCTATCTTTGGCTCCAAAGATGTCAATGTTCAATTTGCAAGCAG 1555
QY 1501 AGCTTGTGATGGGGAACCTTCATCCACCCTGACTGGGATATGTTCCATCCACCACCT 1560
DB 1556 AGCTTGTGATGGGGAACCTTCATCCACCCTGACTGGGATATGTTCCATCCACCACCT 1615
QY 1561 TGTGCGGCTTCCATGCTGCTGCTGAGCCATCTGTGGTGGCCGATATGTTAGTGTAT 1620
DB 1616 TGTGCGGCTTCCATGCTGCTGCTGAGCCATCTGTGGTGGCCGATATGTTAGTGTAT 1675
QY 1621 TCTGTGGGAAGCAATACCTTTGATCTTCTGAAAAAACTAAGTCTTCAGATGATCATC 1680
DB 1676 TCTGTGGGAAGCAATACCTTTGATCTTCTGAAAAAACTAAGTCTTCAGATGATCATC 1735
QY 1681 CTTTGAAGTGAATATGATGATCCGCACTCCGACTGGGATGTTGTTGAAGACCTTTGAT 1740
DB 1736 CTTTGAAGTGAATATGATGATCCGCACTCCGACTGGGATGTTGTTGAAGACCTTTGAT 1795
QY 1741 AATGAGAAACTATGCTTAAGATTTGGAATCTAACAAGTTCACTGAGTGAATGGTGA 1800
DB 1796 AATGAGAAACTATGCTTAAGATTTGGAATCTAACAAGTTCACTGAGTGAATGGTGA 1855
QY 1801 TTTCAACTGCAAGAGAGAGATGTTGTGTGAGACACGCGCAACCAATGCTTTTCACAA 1860
DB 1856 TTTCAACTGCAAGAGAGAGATGTTGTGTGAGACACGCGCAACCAATGCTTTTCACAA 1915
QY 1861 TTACTCAAAACGATGACATCCAAACTAACCCAAAGACATGAGATGGACAGTGGAGAA 1920
DB 1916 TTACTCAAAACGATGACATCCAAACTAACCCAAAGACATGAGATGGACAGTGGAGAA 1975
QY 1921 AACCTATCTCTATTTGAAGGCGCTTAAACCTTTGGCGTTTACCTCTATTCACACAAAAA 1980
DB 1976 AACCTATCTCTATTTGAAGGCGCTTAAACCTTTGGCGTTTACCTCTATTCACACAAAAA 2035
QY 1981 CTTATCTCTCTCAACGCTCTCAAGATCTTGACATAGCTCTTGACCCATTCGAATTCGAG 2040
DB 2036 CTTATCTCTCTCAACGCTCTCAAGATCTTGACATAGCTCTTGACCCATTCGAATTCGAG 2095
QY 2041 CTTCATCAGTGTTCACAGGTGCAACGCAACGTCATCCAACTCTCTACATTTGGCCCAATT 2100
DB 2096 CTTCATCAGTGTTCACAGGTGCAACGCAACGTCATCCAACTCTCTACATTTGGCCCAATT 2155
QY 2101 GGGCTGTGTAACATGCTTAACACTAGTGAAGCCATCCAAATCTGTGAGACTATGACGATGAC 2160

|||||
Db 2156 GGGCTGTGTAACCTGTTAACTAGTGGAGCCATCCAACTGTGATGACGATGAC 2215
QY 2161 CTAACTCAGTCAGATTTGGTGTCAAGAGGTGTGTGAGATGCGAGTATTTGCATGAAA 2220
|||||
Db 2216 CTAACTCAGTCAGATTTGGTGTCAAGAGGTGTGTGAGATGCGAGTATTTGCATGAAA 2275
QY 2221 AAACCAAGGCTTGTGCTATTTGATGGGAGAGATGTTGGTTCAAGTATGATCAGACCAA 2280
|||||
Db 2276 AAACCAAGGCTTGTGCTATTTGATGGGAGAGATGTTGGTTCAAGTATGATCAGACCAA 2335
QY 2281 ATGGTGTGTTCAAGTCCATGGCCATTTGATCTTCATCGGCTGCATTTGGTTATC 2340
|||||
Db 2336 ATGGTGTGTTCAAGTCCATGGCCATTTGATCTTCATCGGCTGCATTTGGTTATC 2395
QY 2341 GAGTACTGTTT 2352
|||||
Db 2396 GAGTACTGTTT 2407

RESULT 4
E36417 2690 bp DNA linear PAT 18-JUN-2001
LOCUS E36417
DEFINITION Raffinose synthase gene.
ACCESSION E36417
VERSION E36417.1 GI:13019215
KEYWORDS JP 2000014389-A/1.
SOURCE Brassica juncea.
ORGANISM Brassica juncea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 2690)
AUTHORS Ejiro, W. and Kenji, O.
TITLE Raffinose synthase gene
JOURNAL Patient: JP 2000014389-A 1 18-JAN-2000:
SUMITOMO CHEM CO LTD
OS Brassica juncea
PN JP 2000014389-A/1
PD 18-JAN-2000
PF 10-DEC-1998 JP 1998351246
PR

PI EJIRO WATANABE, KENJI OEDA
PC C12N15/09, A01H5/00, C12N1/21, C12N5/10, C12N9/88, C12Q1/68// PC
(C12N1/21, C12R1:19), (C12N5/10, C12R1:91), (C12N9/88, C12R1:19), PC
(C12N9/88, C12R1:91), C12N15/00, C12N5/00, C12N5/00, C12R1:91) CC
FH Key location/Qualifiers
FT CDS (134)..(2467).
FEATURES
source location/Qualifiers
1..2690
/organism="Brassica juncea"
/db_xref="taxon:3707"
BASE COUNT 648 a 608 c 668 g 765 t 1 others
ORIGIN

Query Match 44.3%; Score 1041.4; DB 6; Length 2690;
Best Local Similarity 68.3%; Pred. No. 1.2e-299;
Matches 1530; Conservative 0; Mismatches 676; Indels 33; Gaps 5;

Db 390 TCGTAATCTCAGGCAATTCATTCATGACATATTCGTTCAAGGTTTGGTGACATA 449
QY 308 CACACTGGGTGGTTCGAAATGATGGGAGATCTTGAAATCGAGACTCAGATTTGATCTTGG 367
|||||
Db 450 CTCACGGGTGGTTCGAAATGATGGGAGATCTTGAAATCGAGACTCAGATTTGATCTTGG 509
QY 368 AGAATCAGATTTGGTTCGACCGTATGTTTCTTCCGATCGTTGAGAGACCGTTCC 427
|||||
Db 510 AGAATCAGGATCGGTCGCTTATGTTCTTCTGCGCTTCTTGAAGGCTCTTTC 569
QY 428 GAACCTGATTCAGCCCGGGGATGATGACTTTGTCATGTTTGTGTCGAGATGTTGCT 487
|||||
Db 570 GTTCATCTTTTCAAGCTTGGGAGACGATGACGTGGCGGTTTGTGTAATCCGGTCCA 629
QY 488 CGAAGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 547
|||||
Db 630 CCGAGGTGACCGGCTGCGAATTTCTGCAAGTTTGTATGTTCAAGCGGAGACGATCCGT 689
QY 548 TTGACTTGTAAAGAGCGGATGAGATGATGATGATGATGATGATGATGATGATGATGAT 607
|||||
Db 690 TCAAGCTCGTAAGACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 749
QY 608 TGGAGAGAAAGACTCCACAGATATGTTGACAAATTCGTTGGTGTGACGTGGAGCGGT 667
|||||
Db 750 TGGAGAGAAAGACTCCACAGATATGTTGACAAATTCGTTGGTGTGACGTGGAGCGGT 809
QY 668 TTATCTTAACGGTTTCATCCACAGCGGTAATAGAGCGGATGATGATGATGATGATGATGAT 727
|||||
Db 810 TTATCTTAACGGTTTCATCCACAGCGGTAATAGAGCGGATGATGATGATGATGATGATGAT 869
QY 728 GTTGTCTCTCCGTTTATGTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 787
|||||
Db 870 GTTGTCTCTCCGTTTATGTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 929
QY 788 ATCCATTCACCAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 847
|||||
Db 930 ATGCTATTCGATGTTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 989
QY 848 TTTTGAATTTCCAGAGAAATTAACAAATTCGATGATGATGATGATGATGATGATGATGAT 907
|||||
Db 990 TTTTGAATTTCCAGAGAAATTAACAAATTCGATGATGATGATGATGATGATGATGATGAT 1043
QY 908 CCGGAGCGGCGGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 967
|||||
Db 1044 -----AAAACGAGTGGGAGATGAAAGCTTGTGTCAGATGATGAAAGAAATTCCTCA 1097
QY 968 CTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1027
|||||
Db 1098 CCGTTATTTACATCTACGTTGACGCGCTTTCGCGTACTGAGGTGTTCTTCGTCGG 1157
QY 1028 AGGTGCTGCTTGCCTGAGCAGCTGTATTCAGCTGCTGCTTCCACCAGGCGTTCGAGA 1087
|||||
Db 1158 GAGCTCTACTCTCTCCGCTCCCTCAATTTGTCGCGGCAAGGCTTCCGCGGGCTTAAGT 1217
QY 1088 TGACGATGAGATTTGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1147
|||||
Db 1218 TGACGATGAGATTTGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1277
QY 1148 CGAGAGAGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1207
|||||
Db 1278 CGGACATGGCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1337
QY 1208 ACGGTGTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1267
|||||
Db 1338 ACGGCTTTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1397
QY 1268 GAGTGAATTTGGCAAGGATATTAACAAGCAATGACCAATTAATTAATTAATTAATTAAT 1327
|||||
Db 1398 GAGTGAATTTGGCAAGGATATTAACAAGCAATGACCAATTAATTAATTAATTAATTAAT 1457
QY 1328 AAGGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1387
|||||
Db 1458 ACGGTACGCGTTATCGCTAGATGAGACCTGATATGATGATGATGATGATGATGATGATGAT 1517

Oy	1388	AAATATCTCTTGGCGCGTGGTGGATGACTTTTGGTGCACGGACCCCTCTGTGATC	1447
Db	1518	AAGCATCTCTTAGGCTCGTGCGTGGAGTACCTTTTGGTCAGGATCCATTCAGCGCA	1577
Oy	1448	CAAAAGCTAGCTTTTGGCTCCAAAGATGTACATGGTTCATTTGTCCAAAGCAGCTGT	1507
Db	1578	TAAAGCGCATATATTGGCTGCAAAGATGGCCATATGTTCCACTGTGCTCTACACAGTCTTT	1637
Oy	1508	GGATGGGGAATTCATTCACCCCTGACTGGGATATGTTCCATCCACCCAGCCCTGTGGCG	1567
Db	1638	GGATGGGAATTCATTCACCCCTGATTTGGGACATGTTTCAGTCCACACATCCCTGTGGCTG	1697
Oy	1568	CTTTCATGCTGCGCTCTCAGAGCATCTCTGTTGGGCCCATCTATGTTAGTATCTGTGG	1627
Db	1698	AGTTTCATGCTGCTCTCTCGGCCATCTCCGGGGGGCCATTTTACATCAGCAATGTGTGG	1757
Oy	1628	GAAGCATTAATCTTGTGATCTCTGTAAATAAACTAGTGCTTCTGATGGATGATCTTCGAA	1687
Db	1758	GCCAGCAGCATTTTCGATCTCTTTGAACGACTGTCTGTGCTTGACGGTTTCGATTTTGAAGT	1817
Oy	1688	GTGAGTACTATGCACCTCCGACTCGCGATTTGTTTGTTCGAAGACCCCTTGGCATATGAG	1747
Db	1818	GTGAGCACTATGCACCTCCCAACTCGTAGCCGTCTCTTTGAACCCCTCTCATGATGACA	1877
Oy	1748	AAATATGCTTAAAGATTTTGGAAATCTCAACAAGTTCACGTGAGATGATTTGTCATTTAACT	1807
Db	1878	AAACCATGCTCTCAAGATTTTGGAACTTCAACAAAGTACCTGGAATTTTGTGAGCATTTCACT	1937
Oy	1808	GCCAAAGGAGAGATGGTGTGCTGAGAACACGCGCCAAACAAATGCTTTTACAAATACTCA	1867
Db	1938	GCCAAAGGAGAGATGGTGTGAGAGAAACCCGACGACCAATGCTTTCTCCAAATGCGTTA	1997
Oy	1868	AACGAGTGCATCCAAAACCTAACCCAAAAGACATAGAAATGCGACAGTSGAGAAAACCTA	1927
Db	1998	ACAGGTTAACCGGCACAAACAAATCTTAAGAGCTTGAATGGAACAGTGGAAACCAACA	2057
Oy	1928	TCTCTATTGAAGCGCTTAAACCTTTTGCGCTTTACTCTATCAACCCAAAACCTATACC	1987
Db	2058	TCTCTCGTTGAAAGCGTTGAAGAGTGTCTTGTGTTCTTGTGCTTAAGTCAAGACCTTGTGT	2117
Oy	1988	TCTCCAAGCCCTCTCAAGATCTTGCATATAGCTTGTACCCATTCGAAATTCGAGCTCATCA	2047
Db	2118	TGCTCGAACCCAAACGATGATCTCGAATCATCTTTGGAGCTTTCAAGTTTGAGCTATCA	2177
Oy	2048	CTGTTTCAGCAGTGCACCAACATCATCAAACTTCTCTACACTTTTGCCCAATTGGCGTGG	2107
Db	2178	CTGTCTCAACAGTTGTTCATATTAAGAGGTAGTTTCGGTTTCAGTTTCTCTCAATCGGATGG	2237
Oy	2108	TGAACATGCTTAAACACTAGTGTGAGGCCATTCATCTGTGACTATATAGCATAGCTTAAGCT	2167
Db	2238	TTAAACATGCTTAAACACTAGTGTGAGGTGCAATTCGATCCCTTGGTGTATCATAGGA-----AT	2291
Oy	2168	CAGTGCAGATTTGGTGAAGGGGTGTGGTGAAGTGTGGAATTTTGCATGCAAGAAAACCA	2227
Db	2292	CCGTTGAGATTTGGAAGTTCTGTGTGCTGTGGAAGTTTCAAGGTTTATCTATCAAGAAACCTG	2351
Oy	2228	GGCGTTGTCTGATTTGATGGGAGAGATTTGGGTTCAAGATGATGACAGCAACCAATAGTGTG	2287
Db	2352	CGAGCTGCAAAATATTGATGTGAAGTGTGTGAGTTTGAGTATAGC---AAGAGTCAATAGTGA	2408
Oy	2288	TGCTTCAAGTGCATGGCC	2306
Db	2409	TGCTTCAAGTGCCTTGCTC	2427

RESULT 5	ARI82382	2498 bp	DNA	1 linear	PAT 20-APR-2002
LOCUS	ARI82382				
DEFINITION	Sequence 23 from patent US 6337430.				
ACCESSION	ARI82382				
VERSION	ARI82382.1	GI:20225298			
KEYWORDS					

[illegible]

QY	57	TCACATGTCGTCACCGTTGGCAATCGACGAGTCGGATTTCACATCTGAAAGCGTCAATTGCTT	116
Db	118	TAATTTCGCTTTGTTCGCAATTAACCCATAGAAGATCAAAATTTCTCTGCGCAAGGCCACCTTT	177
QY	117	TCCTGTCGAGTCTCCCTGGAACATTGTTCCTTCCTTCCTCGATC-----AC	164
Db	178	TCTCAGCGGAAGTCTCCCGAAACAATAATAGTCAACCCCTTTCACCCCTTCACCCCTCAAGAGTAG	237
QY	155	TTTCGATTACAAGATGCCCGGTTTGCGTTTGCTTTGTTGGATTTCGACGCCCTCGAACC	224
Db	238	TAAAGAACACAGAGGACGACGACGTCGTAGTGTTGGTTGCTTGCGGCTTCACGCCGACGAGCC	297
QY	225	TGATAGCCAGCATGTTGTTTGCATTGGGAAGCTCAAGAGATATTCGGTTTATGAGATATTTT	284
Db	298	CAGAAAGCCGACACGCTGGCTTCCCTGGGGAAGCTCAGAGGAATTAATAATTCTATGACCATTT	357
QY	285	CAGGTTTAAAGCTTTTGCTGGGACTACACACTGGGTTGGTTCGAAATGCTGGGAGCTTTGAATC	344
Db	358	CCGGTTTAAAGTGTGGTGAGACCATCTACCTGGGTGGTAGAACACGACACAACTGAGCA	417
QY	345	GGAGACTCAGATTTGATGATCCCTTGGAAGTACG---ATTCTGGTCGACCGTATGTTTCCCT	401
Db	418	CGAGCACACAGATGATGCTTCTCGACAAAAACGACACACTCGGACGCCCTTTTGTTGGTAT	477
QY	402	TCTTTCGATTCGTTGAGGAGACGTTTCCGAACTCGCATTCAGGCTGGGATGATGACTTTGT	461
Db	478	TCTCCGATTCCTCCAAAGCGCTGTTCCAGAGCCTCCCTCGACCCGGTTTGGATGATTACGT	537
QY	462	CGATGTTTGTGTCGAGAGTGTCTGTCGAAGATTGATGATCATGTTCCGGAAGTATGTT	521
Db	538	GGACGCTTTGCATGGAGACGGGGTGAGACACGTCCTGTGGCTCCAGCTTCGGGAGCTCTTT	597
QY	522	GTATCTTCATCTGGTGATGATCCGTTTGGCATCTGTTAAGAAGGCGATGAAGATCTGGAG	581
Db	598	ATACGTCACAGCTGGCCCAATGACCCGTTATCACTTCCTTAAGAACCAACTAAAGTCGTTAG	657
QY	582	GACCCATCTTGGAACTTTTCCGTTGTTGGAGAGAAACACTCACACAGTATCTCGGACAA	641
Db	658	GATGCATTTTGGGAGCGTTCAAGCTTCTCGAGAGAGAAAAACCGCGCATGATCATAGACAA	717
QY	642	ATTGCGTTTGTGTCACGTCGGGACCGGTTTTACCTAACGTTTCATCCACAGGGCGTAATAGA	701
Db	718	GTTTGTGTTGTGTACATGAGGACGGGTTTTACTTGAAGTGTCATCCCAAGTGTGTGGGA	777
QY	702	AGGCGTGAAGGATCTGCTCGACGGCGGTTGTGTCGCCCGGTTTACTCTTAATCGACGATGG	761
Db	778	AGGGGTGAAGGATTGGTGGAGGAGAGGGTGGCTTCACAGGGATGGTCTTAATTCGACGACGG	837
QY	762	TTGGCAATTCATCGGACAGATTTCGGATCCCATAC---CAAAAGAGAAATGAACCAAC	818
Db	838	GTGGCAACCATTTTGTCTACGACGAGAGAACCCCATTAACGAGCAAAAGGGTATGAAGCAAC	897
QY	819	CGTCCGCGGACGCAAAATGCCCTGCTTTTGAATTTCCACAGAGAAATTACAATTTCCG	878
Db	898	CTCCGACAGGGGAGCAAAATGCCATCAGGTTGTGTAAGTTGGAGGAAAAATTACAAGTTACG	957

QY	879	TGCTACGTCATATCCCAAGGCCACCGGCCCCGACGGCCGACGAAAGGGATATACAGCGTT	938
Db	958	ACAGATATTGTAGTGGCAAGGATTTCTG-----AGAAAGGATATGGGTGCCCTT	10020
QY	939	TATAGATGACACTCAAGAGAGATTTTAAAGCTGTGGAGCATGTTTATATGTTTGGCATGCTTT	998
Db	1003	TGTTTAGGGACTGTGAAGCAACAGTTTATAGAACCGTGGAGCAGAGTGTATGTGTGGCACCGCGCT	10678
QY	999	GTTGTGATATTTGGGGTGGCCCTTGCCCGCGAGTGGCTTGCCCTGAGGACAGCTGTGAT	10588
Db	1063	TTGTGGGATATTTGGGGTGGGGGTTCAGACCCAAAGTGTCCGGGCAATGCCCCAGCGTTAAGTTGT	112228
QY	1059	TCACACGAGTGGTTTCACACAGGCGTCGAGATGACAGATGGAGAGATTTGGCGGTGGATAAAT	11188
Db	1123	CACGTCCAGAGCTGTCCATATGGACTATAAATTGACAAATGAAGAGATTTTAGCGTGTGATMAAT	11820
QY	1119	TGTTCTTCATTAAGGTGCGGGTGGGTCGCCGCGGAGAGAGGCTGAGAGAGATATACGAGAGCT	11788
Db	1183	CGTCAAGTACGGAATTTGGACTGTGTGCCACACACCTGGCTCACCTTTTGTACGAGGGCGCT	12424
QY	1179	TCATGCTCATTTTGGAAAAAGTTGGGATTCGACGGTGTTAAGATTTGACGTTATCCACCTATT	12388
Db	1243	CCACTCCCGTTTGGAAATCTCGGGGTATTTGACGGTGTTAAGTTTGAAGTTATACACTTCT	13020
QY	1229	GGAATATGTTGTGTGAGACATATGGAGGAGAGTGGATTTGGCAAAAGCATATTACAAAGC	12988
Db	1303	CGAATATCTATCCGAGGAATAACGCTGGCGGTGTGAGCTAGCCAAAGCTTATTACAAAGC	13620
QY	1299	AATGACCAAAATCAATAATTAACATTTTAAAGAAATGGAGTCATTTGCAAGATATGACAA	13588
Db	1363	GCTACATGCTTCGGTGAAGAGCATTTTCAAAGGCATATGGGGTCAATTTGGAGCATGGACAA	14222
QY	1359	TTTGTAACGACTTCATGTTCTCTTGGACAGAGAAAGCTATCTCTTGGTCGTGTTGTGATGA	14188
Db	1423	TTGTAAATGACTTCTTCTCTCTGTAACCGAAGCATAGCCCTTGGGGCGGTGTAGAGATGA	1482
QY	1419	CTTTTGGTGCAGCGAGCCCTCTGTGATATCCAAAGGTACGTTTGGTCTCAAGATGTCA	14788
Db	1483	TTTTTGGTGCAGTATCCCTCTGTGAATATCCAAATGGCACGTATTTGGCTCTCAAGGGCTGTCA	1542
QY	1479	CATGCTTCATATTGTGCACACACAGCTTGTGATGGGGAACCTTCAATCCACCCGACTGTGGA	15388
Db	1543	CATGCTGCACCTGTGCTTACACAAAGCTTTGTGATGGGGAATTTATTTCACCCGAGATTGGGA	1602
QY	1539	TATGTTCCAAATCCACCCACCCCTTGTGCGCGCTTCCATGCTGCTCTCGAAGCCATCTGCG	15988
Db	1603	CATGTTCCAGTCCACATCCACCTCTGTGCCCAATTCATGACGCGCTTAAGGGCCATCTGCG	1662
QY	1599	TGGCGCCATCTATGTATAGTATCTCTGTGGAAAGCATTAACCTTGTATCTTGTGAAAAACT	16588
Db	1663	TGGACCAAGTTTACGTTATAGTATGTTTGTGAAAGCAACAATTTCAAGTTGCTCTCAAGAGCT	1722
QY	1659	AGTCTCTCTGATGATGACATCCCTGTGAGTGAATGATGTGACATCCGACCCGACCGGATGT	1718
Db	1723	CGCTTGCCTGATGGAGCAGATTTTGGCTTGTCACTATGTGACTTCCACACGAGACTG	1782
QY	1719	TTTTTGTGAAGACCCCTTGTGATATAGAGAAACTATGCTTAAGATTGGAAATCTCAACA	1778
Db	1783	TTTTTGTGAAGACCCCTGTGATATGAGAAAGCAANTGCTCAAAATTTTGGAAATCTCAACA	1842
QY	1779	GTTTCACTGAGATATTTGTGATTCATTAACCTGCCAAGAGAGAGATGCTGTGAGACAGC	1838
Db	1843	ATATATACAGGTGTTTGGGTCTATTTTAAATTGCCAAGAGAGTGGGTGTCTCCCGTATAG	1902
QY	1839	CCGGAACCATGCTTTTCAACAATATCCAAACAGAGTGAATATCCAAACTAACCCAAAAGA	1898
Db	1903	GAGAAACAAAGTGCCTGTGAATTTTCAAAAATGTGACATGCTTAAGAGATCTCTCAAGA	1962
QY	1899	CATGATGATGACAGTGGAGAAAAACCTTCTCTATTTGAGAGCGTTAAACCTTTGGCGCT	1958
Db	1963	CATTGAATGAGCAATGGGAAAAAGCCCAATATGACATAAAAGGGATGAATGTGTTGCTCT	2022
QY	1959	TTTACCTCTATCAAGCCAAAAAATTATCTCTCAAGGCCCTCTCAAGATCTTGTGATAGC	2018

Db	2023	ATATTGTTCAAGGACCAACAACTAAAGCTCATGAGGCATCGAGAAATTGGAGCTTC	2082
Qy	2019	TCCTGACCATTGCAATTGCGAGCTCATGCTACTGTTTCACACAGTGACCAATCATCCAAAC	2078
Db	2083	ACTTAGGCATTACTCTTTGAGCTATTTGACASTGTCTCCAGTAGANTGTGCTCAAAAAA	2142
Qy	2079	TTCCTACACTTTGGCCCCAATTGGCTGGTGACATGCTTAACTAGTAGGACCATCCA	2138
Db	2143	GTTAAATTCAATTGCTCCCAATGATGTAGTGAACATGCTTAACTGGTGGTGCATTTCA	2202
Qy	2139	ATCTGTGACCTATGACGATGACCTAAGCTCAAGTCAGATTGGTGTCAAAAGGCTGGTGA	2198
Db	2203	GTCCATGGAGGTTTGACAAACCATATGANTGTGCTAAATTTGGGTTGTAGGGTTGTGGGGA	2262
Qy	2199	GATGCGAGATTTTGGATCGCAAAAACCAAGGCTGTGCTGATTTGATGGGAGAGATTGG	2258
Db	2263	GATGAAGGTTGTTGTCATCAGAGAAACCAAGTTAGTTGCAACTAGATGGGTTAGTTGTAA	2322
Qy	2259	GTTCAAGTATGATCAGACGACCAANTGGTGGTTCAGTGCATGCGCCAAATTGATTCTTC	2318
Db	2323	ATTGATATTATTTT---AGSATTAATAATCTGAGATGTCAGAACTTCCCTGAGCTTCAAA	2379
Qy	2319	ATCG 2322	
Db	2380	ATTG 2383	
RESULT 6	E37133	2690 bp	DNA linear PAT 31-JAN-2002
LOCUS	E37133	2690 bp	DNA linear PAT 31-JAN-2002
DEFINITION	Ratifinose synthase gene.		
ACCESSION	E37133		
VERSION	E37133.1 GI:18626585		
KEYWORDS	JP 2000166562-A/1.		
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 2690)		
AUTHORS	Watanabe, E. and Oeda, K.		
TITLE	Ratifinose synthase gene		
JOURNAL	Patent: JP 2000166562-A 1 20-JUN-2000;		
COMMENT	SDMITOMO CHEM CO LTD		
	OS Beta vulgaris L.		
	PN JP 2000166562-A/1		
	PD 20-JUN-2000		
	PF 04-DEC-1998 JP 1998345590		
	PR		
	PI EUJIRO WATANABE, KENJI OEDA		
	PC C12N15/09, C12N1/21, C12N5/10, C12N9/88, C13Q1/68//C12N15/09, PC		
	C12R1:91),		
	PC (C12N9/88, C12R1:91), C12N15/00, C12N5/00, C12N15/00, C12R1:91) CC		
FEATURES	source		
	FT CDS	Location/Qualifiers	
		1..2690	
		/organism="unidentified"	
		/db_xref="taxon:32644"	
BASE COUNT	762 a 507 c 641 g 778 t	2 others	
ORIGIN			
Query Match	41.5%; Score 975.4; DB 6; Length 2690;		
Best Local Similarity	67.8%; Pred. No. 6.8e-280;		
Matches 1443; Conservative 1; Mismatches 662; Indels 21; Gaps 5;			
Qy	183	GGTTTCGGTTGGTTCCTTTGTTGATTCGACGCTCGGACCTGATAGCCGACATGTTGT	242
Db	403	GGCTTTTATATGGGCTTTTGTGGTTTGAATGTCGACAGCCCAAGGCCCGCACGTTGT	462
Qy	243	TTCGATTGGGAAGCTGAAGATATTCGGTTATGATGATATTTTCAAGTTTAAGTTTGGTG	302
Db	463	ATCCGTGGGCGACGATTCAGGAATTCCTTCATGATATTCCTTCAAGGTTCAAGGATGTTGGT	522

OY	303	GACTACACACAGTGGGTGCTCGAANAATGGTGGAATCTTGAAICGAGACTCAGATTGTGAT	362
Db	523	GACTAACCCATTGGACGTGGGTCGCMAATGGGGGAGCACTTGAGCATGAGACCACAATTTCTCAT	582
OY	363	CCTTAGAACATCAGA-----TTCTGCTGTGACCGTAATGTTTTCCATTCCGATCGTTGA	416
Db	583	CCTTGATAAGTCAGATGAMAGCTTTGGGCCGCTCCATATATTGGATCCCTCCATTGATCGA	642
OY	417	GGGACGCTCCGAACCTCGATTAGCCTGG--GATGATGACTTTGTCGATGTTTGTGT	473
Db	643	AGGCCCATTTTCGGGCATCTCTCCAGCCGGSTCTGTGATGACTATGNGAATATGTGT	702
OY	474	CGAAGTGTCTTCGCAAAATTTGTATGATCATCTGTTCCGAATATGTTGTATCTTCAATGC	533
Db	703	TGBAAGTGGGTCCACTMAAATTTGTCGGAAATCTGTTCCGGCTGTTCTTATATACGGGC	762
OY	534	TGTGATGATCTCGTTTGTCATCTGTTTAAAGAGCGATGATGCTGAGAGCCCATCTTGG	593
Db	763	TGGGCCCTACCCCATTTTAAGTTATTTAAAGATACAAATGAMAGAAAGTCCAAAGCCANTTAG	822
OY	594	AACTTTTCCTTGTGGAGAGAAAGCTCACACAGGTATCGTGACAAATTCGTTGGTG	653
Db	823	GACTTTCAAACCTCTTAGATATACAAAATCTCCAGGAATAGTGCAAGATTGGATGGTG	882
OY	654	CACGTGGGAGCGCTTTTAACTAACGGTTTCACTACAGGGCCGTAATAGAAAGCGTGAAGCA	713
Db	883	TACATGGGATGCTATTTTAACTCAAAATGAGCCMTATGGTGTGGGAAGGATTAAGG	942
OY	714	TCTCTGTGACGAGCGCGTGTGTCCTCCCGTTTATGCTCAATCGACGATGGTTGGCATCCAT	773
Db	943	ACTGTGTGAAMAAGCGGGTCCACCGSGTCTGACTCATTTGATGATGGGTGGCATCTAT	1002
OY	774	CGGACACGATTCGGATTCCTACACA---AAGAAGATAGAACCAACCGTCGCCGGCGA	830
Db	1003	TTTGATATGACGATGATCTCGATTACCGACAGAAAGGAGATTAACCGGACTTCTGCCGGGGA	1062
OY	831	GCAAAATCCCCGCGCTTTTGAATTCACAAGAAATTCAAATTCCTGTGATACGTCA	890
Db	1063	GCAAAATGCATGTGANTGATGATCAAGTACAGAGAAACCTTCAAGTTTAAAGCACTATMAAG	1122
OY	891	TTCCAAGGCCACGGGCCCGGAG---CGGGCCAGAAGGGGATGAAGCGCTTTATATGATGA	947
Db	1123	CCCAAAATATTATGGGCCCATGAGGATCATCCCAATATGGAAGAAAGAGGCGCTTGTATAGGA	1182
OY	948	ACCTAAAGAGAGATTTAAAGCTGTGGAGCATGTTTATGTTGGATCGCTTGTGTGATGA	1007
Db	1183	CCTTAAAGSAGAGTTCAAAACTGTTGAGCATGTGTATTTTGGCATGCTTTTACGGGCTA	1242
OY	1008	TTTGGGTGCCCCCTTCGCCGAGGTGCTCTGCTTCCCTGAAGGACGTGTGATTCAGCCAGT	1067
Db	1243	TTTGGGAGGGGTAAAGGCCCAATGTTTCCAGGCTTACRGAGGCCCAAGTATGTAACCCANA	1302
OY	1068	GCTTTTACACAGGGCTGCAAGTATGACATGAGAGATTGGCGGTGATTAAGATTGTCTTCA	1127
Db	1303	GCTTTTCCCCTGGGTCTTGAGATATGACAAATGGAAGATTCTAGCTGTGGATTAATAATTTGTAA	1362
OY	1128	TAAAGTGGGCTGTCTCCGCCGAGAGAGGCTGAGAGATGTACGAAGAACTTCAATGCTCA	1187
Db	1363	TGTGATTGGCTGTCTCAGGCTGATTAAGGCCCAAGAACCTTTATGAAGGGTTGCATTTCA	1422
OY	1188	TTTGGAAAACTTGGATTCACACGCTGTTTAAAGATTGACGTTATTCACATTTTGGAGATGTT	1247
Db	1423	TTTGGAAAAATTGTGGATTATGATGAGTCAAAAGTTGATGTCATCTCAATTTGTTGGAGATAT	1482
OY	1248	GTTGGAAGACTATGAGAGGAGAGTGGATTTGGCAAAAGGATATTTTCAAAAGCAATGACAA	1307
Db	1483	GGCAGAGGACTATGAGAGAAAGATGTGAACAGAAAACATCTATTAAGGCATATACAGA	1542
OY	1308	ATCAATAATAATACATTTTAAAGGAAATGAGTCAATGCAAGTATGAAACATTTGTAACA	1367
Db	1543	ATTCAGTGGTATGACATTTTCAAGGCAACGCTGTGATTTGTACATGAGAGCACTGCAACA	1602

QY	1368	CTTCAATGTTCCCTTGGCACGGAAGCATCTCTCTTGGTGCTGTGTGTGATATGACTTTTGGTC	1427
Db	1603	TTTTCATGCTCTTGGTACTGAGACCATTGTGCTGTGGCTGGGAGTGACTTTTGGCC	1662
QY	1428	CACGGACCCCTCTGGTGAATCCAAAGGTAAGCTTTTGGCTCCAAAGGATATGCATGGTTCA	1487
Db	1663	AACATATCCGCTTGGAGATATTAATATGTAACATATTGGCTCCAAAGGCTGTCAATTATGTGCA	1722
QY	1488	TTTGTGCCAAGCAGACACTTGTGTGATGGGGAAGCTTCAATCCACCCTGACTGGGATATGTGCCA	1547
Db	1723	TTGTGCGCTAACATTAAGCTTATGGAATGGGAAAGCTTTATACACCCTGACTGGAGCATGTGTCCA	1782
QY	1548	ATCCACCACACCTTGTGGCGGCTTCCATATGCTGCTCTGTGAGCCATCTTGTGTGGCCGAT	1607
Db	1783	ATCTACACACCCCTTGTGGTGAATTTTCATGCTGCATCTGTGGATTTCTGTGTGGACCAAT	1842
QY	1608	CTATGTAGTGAATTTGTGGGGAACCATTAACCTTATCTCTGAAAAAATAGTGCTTCC	1667
Db	1843	TTTATGTTAGTGAATGTTGTTGGCAACATTAACATCCCTTGCTCCAAAGGCTGTCTTGTGGC	1902
QY	1668	TGATGATCGATCTCTTGCAGAGTGAATAGTACATCTCCGACTGCGATGTTGTGTTTGA	1727
Db	1903	TGATGGTTGATCTCTTGCATGGAGTACACATGCACTTCTCTACTAAGATTTGCTATTGTTGT	1962
QY	1728	AGACCCCTTGCATTAATGAGAACTATGCTTAAGATTTTGGAAATCTCAACAGTTCACTGG	1787
Db	1963	AGATCCCTTGTGCAGATGGCCAAACAAATGCTCAAAATTTTGGAAACCTCAACAGTACAAATG	2022
QY	1788	AGTGAATGGTGCATTCACACTGCCAAGGAGAGATGTTGTGTGAGACACCCGCAACCA	1847
Db	2023	AGTGTCTGGAGTCTTCAATTGCCAAAGGAGGAGGTTGAGCCGTGCTGTCCAAAAAATCT	2082
QY	1848	ATGCTTTTCCAAATCTCTCAAAAACGAGTACATCTCAAAACTAACCCAAAGACATTAAGATG	1907
Db	2083	ATGTTTCTCAGAGTATTCAAAACCTATTTCTCGCAAGACAAAGTCCAAAAAGATGTTGAATG	2142
QY	1908	GCACAGTGAAGAAAACCTATCTCTAATTGAGAGCGTTAAAACCTTTGCGCTTAACTCTTA	1967
Db	2143	GGAGAACGACACCAAGCCATTCCTCCATCAAAAGAGATGGAATGTTTGGCATGATCTTCAC	2202
QY	1968	TCAGAGCAAAAAACATTAATCCCTCTCCAAAGCCCTCTCAAGATGTTGATCAATAGCTCTGAGCC	2027
Db	2203	CAGGAAAAAAAGCTAATCTCTCTTCACAACTATCTACACCATTTGAATAATCACTTGAATCC	2262
QY	2028	ATTGCAATTCGAAGCTCATCACTGTTTTCACAGTGAACCAAACTCATCAAACTTCTCTACA	2087
Db	2263	CTTTCGATTACGAGCTTATTGTAGTCTCTCCGATGACAAATTTCTAACCTGGAGATGCAATGCC	2322
QY	2088	CTTTGGCCCCAATTTGGCTGTGTGACATGCTTAAACACTAGTGGAGCCATTCATCTGTGGA	2147
Db	2323	ATTGTACCCCAATAGGATTAAGTAACATGCTCAACCCCGAGGCGCATGTCAAATGCTTTTGA	2382
QY	2148	C-----TATGACGATGAGCTTAAGCTCAAGTGCAGATTTGGTCCAAAGGGTGTGTGSAAT	2201
Db	2383	CATCACTGAGGATTAATGAGGATTAAGATGGTTCAGGTTGGTATTAAGAGGCCCGGAGAAAT	2442
QY	2202	GGAGATTTTTCATGGAAAAAAACCAAGGGCTTGTGTAATGTTGGGAGAGATGTTGGGTT	2261
Db	2443	GATGGTTTATTCATCAGAAAACCAAAAGCCGTGTAGATTAATGAGAAAGACATGGAATTT	2502
QY	2262	CAAGTATGATCAAGACCAATGGTGT 2288	
Db	2503	TGAGTATGAGAGACATGATTAAGTT 2529	

RESULT 7					
LOCUS	BD010263				
DEFINITION	BD010263	2343 bp	DNA		
ACCESSION	BD010263	Mutant protein, its DNA	and use thereof.		
VERSION	BD010263.1	GI:18638636			
KEYWORDS	JP 2001078783-A/1.				
SOURCE	Glycine max.				
					PAT 31-JAN-2002
					linear

ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 2343)
Watanabe, E. and Oeda, K.
Mutant protein, its DNA and use thereof
Patent: JP 2001078783-A 1 27-MAR-2001;
JOURNAL SUMITOMO CHEMICAL CO LTD
COMMENT OS Glycine max (soybean)
PN JP 2001078783-A/1
PD 27-MAR-2001
PF 03-JUL-2000 JP 2000200571
PR
PI EJIRI MATSUNABE, KENJI OEDA
PC C12N15/09,A01H5/00,C12N5/10,C12N9/10,C12Q1/68,C12N15/00, PC
C12N5/00
CC
FH Key Location/Qualifiers
FT CDS (1)..(2343).
FEATURES
source 1..2343
BASE COUNT 597 a 493 c 648 g 605 t
ORIGIN
Query Match 41.1%; Score 966.8; DB 6; Length 2343;
Best Local Similarity 66.0%; Pred. No. 2.5e-277;
Matches 1508; Conservative 0; Mismatches 737; Indels 39; Gaps 6;
57 TGACATGTCGTACCGCTTTCATTCAGACGAGTGGATTCTACTGTGACGGTCTT 116
117 TCTGTCGATGTTCTCGAGACATTGTTGCTTCTCTTCGCTAC-----AC 164
117 TCTCAGGAGATTCCTCGAATAATAGTACCCCTTCACCCATCGACGCCAAGAGTAG 176
165 TTGATAGACAAATGCCCGTTTGGTGGTCTTGTGGATTGATGAGCGCGGACACC 224
177 TTAAGAACAGAGAGACGACGACGCTCGTAGGTGCTTGCGGCTTCACCGGACGAGCC 236
225 TGTATCCGACGATGTTGTTGATTCGATTGGAAAGCTGAAGCATATTCGGTTTATGATATTT 284
237 CAGAACCCGACAGTGGGCTTCCTCGGGAAGCTCAGAGAGATTAATTCATGAGCATATT 296
285 CAGGTTTAAGGTTGGTGGACTACACACTGGGTTGTCGAATGTGGGGATCTTGAATC 344
297 CCGGTTTAAGGTTGGTGGAGCACTCAGTGGGTCGTTAGCAGACGACGAACTGGAGCA 356
345 GGAGACTCGATGTGATCCTTGAAGAGTCAG---ATTGCTGTCGACCGCATGTTTCC 401
357 CGAGACACAGATGATCCTCTGACAAAAACGACGACGCTCGAGCGCCCTTGTGTTGAT 416
402 TCTTCCGATGTTGAAGGACCGTTCCGAACTCGATTGAGCCCTGGGGATGATGATTTGG 461
417 TCTCCGATCCTCCAGGCTCGTTCCGAGCTCCCTCCGCAACCGGTTTGGATATTACG 476
462 CGATGTTTGTGTCGAGAGTGGTTCGTGAAAGTTGTTGATGATCGTCCGAAGTATGTT 521
477 GGACGTTTCAATGAGAGGGGCTCGACAGCTCTCTGCTCAGGTCGGGAGCTGCTT 536
522 GTATCTTCATGCTGATGATGATCGTTTGCACCTGTTTAAAGAGCGCATGAAGTCTGAG 581
537 ATACGTCACGTTGGCATTGACCCGATACAGTTGCTTAAAGAAAGCACTAAAGCTCTTAA 596
582 GACCACTCTTGAACCTTTGCTGTTGGAGAGAGACTCCACAGATATCGTGGACAA 641
597 GATGCTTTGGGAGCTTTCAAGCTTCTCGAGAGAAACCGGCCAGTATGTCACAA 656
642 ATTGGGTTGTGACAGCTGGAGCGCTTTTACCTAACGCTTACATCCAGGGCGTAAATAGA 701

1177 TCTGCTCATTTGGAAAAAGTTGGATCGACGCTGTTAAGATTGACGTTATCCACTATT 1238
1179 CCACCTCCGTTTGGAAATCTCGCGGATTTGACGGTGTAAAGTTAGACTTATACACTTGC 1238
1239 GGAGATTGTTGTGGAAGACTATGAGGAGAGTGGATTTGGCAAGGCATATTACAAAG 1298
1239 CGAGATGCTATCCGAGGATATCGGTGCGCTGTTGACATACCAAGACTTATTACAAAG 1298
1299 AATGACCAATCAATTAATTAATTAAGAAATGAGATTCATTGCAAGTATGAGACA 1358
1299 GCTCAGTGGTGGTGAAGAGCATTTCAAGGCATTTGGGTCATTGGCAGCATGGAGCA 1358
1359 TTGTAACGACTTATGTTCTTGGCAAGAGTATCTCTTTGGTCGTTGGTGTATGA 1418
1359 TTGTAATGACTTCTCTCTGCTGACCGAAGCCATAGCCCTTGGCGCGGTAGAGATGA 1418
1419 CTTTGTGTCAGGAGACCGCTGCTGATCCAAAGGATGATGTTTGGCTCCAAAGATGCTA 1478
1419 TTTTGTGTCAGTCAATCCCTTGTGAGATCCAAATGACAGTATTTGGCTCCAAAGGATGCTA 1478
1479 CATGTTTCATTTGTGCAACGACGCTTGTGATGGGAACTTCATTCACACCTGACTGGGA 1538
1479 CAGTGTCACGTCGCTTACAAACAGCTTGTGATGGGAAATTTATTACAGCGATTGGGA 1538
1539 TATGTTCAATCACCACCGCTTGTGCGCTTCATGCTGCTCGAGCCATCTCTGG 1598
1539 CATGTTTCAGTCCACACCGCTTGTGCGCTTCGAAATTCATGAGCCTTCAAGGCGCATCTCTGG 1598
1599 TGGCCGATCTATGATGATGATCTGGAAGAAAGCACTAATCTTGTGATCTTGAAGAAACT 1658
1599 TGGACCAATTTACGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1658
1659 AGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1718
1659 CGCTTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1718
1719 TTTGTTGAAGACCTTTGCAATATGAGAAACTATGCTTAAAGATTGGAATCTCAACA 1778

Db	1719	TTTGGTTGAAGACCCCTTGCATGATGGGAGACAACTGCTCAAAATTTTGGAAATCTCAACAA	1778
Db	1779	GTTCACTGGAGTGGATTTGGTGCATTTCACTGCCAAGAGAGAGATGCTGCTGAGACACG	1838
Db	1779	ATATTACAGTGTGTTTGGGCTCTATTAAATTTGGCCAAAGAGAGTGGGTGTCGCCGTACTAG	1838
Db	1839	CCGCAACCAATTCCTTTTTCACATFACCTCAAAACGAGTGCATCCAAACCTCAACCCAAAAGA	1898
Db	1839	GAGAAACAAAGATGCTCTGCAATTTTTCACAAACCTGTGACATGCTTAGCGAGTCTCAAGA	1898
Db	1899	CATAGAAATGGCAGCATGGAGAGAAACCTATCTATTGAGAGCGTTTAAACCTTTGGCGT	1958
Db	1899	CATTGAATGACACATGGGAAAGAGCCCAATATGCACTAAAGGATGAATGTGTTGCTGT	1958
Db	1959	TTACCTCTATCAAGCCAAAACCTTATCTCTTCACAGCCCTCTCAAGATCTTGACATAGC	2018
Db	2019	TCTTACACCATTCGAAATTCGAGCTCATCTGTTTTCACCAATGACCAAACTATCCAAAC	2078
Db	2019	ACTTATGACCATTTACTTTTGGACCTTTTGACATGCTCTCCAGTGAATGTGCTCAAAA	2078
Db	2079	TTCTCTACACTTTGGCCCAATGGAGCTGGTGAACATGCTTAACACTAGAGCCCATCA	2138
Db	2079	GTTAATTTCAATTTGCTTCCATATGATGTAGTGAACATGCTTAACACTGGTGGTGCATTTCA	2138
Db	2139	GTCATGAGAGTTTGAACACACATAGATGTGTCAAAATTTGGGGTTAGGGGTTAGGGGA	2198
Db	2199	GATGCGAGATTTTGCATCGAAAAACCAAGGCTTGTCTATTGATGGGAGAGATGTTGG	2258
Db	2199	GATGAGAGGTTTGGATCAGAGAAACCAAGTTAGTTGCAAACTAGATGGGGTATGTTGAAA	2258
Db	2259	GTTCAAGTATGATCGAGCAACCAATGCTGTGTGTGTTGCAAGTGCATGATGATTTCTTC	2318
Db	2259	ATTGATTAATG---AGGATTAATGCTGAGAGTGCAGATTTCCCTGGCCTAGTGTTCAAA	2315
Db	2319	ATCG 2322	
Db	2316	ATTG 2319	
RESULT 8			
E24424			
LOCUS	E24424	2497 bp	DNA
DEFINITION	Raffinose synthase gene and use thereof.		linear
ACCESSION	E24424		PAT 18-JUN-2001
VERSION	E24424.1	GI:13018164	
KEYWORDS	JP 1999215984-A/2.		
SOURCE	Glycine max.		
ORGANISM	Glycine max.		
	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.		
REFERENCE	1 (bases 1 to 2497)		
AUTHORS	Ejilro, W. and Kenji, O.		
TITLE	Raffinose synthase gene and use thereof		
JOURNAL	Patent: JP 1999215984-A 2 10-AUG-1999;		
	SUMITOMO CHEM CO LTD		
	OS Glycine max(soybean)		
	PN JP 1999215984-A/2		
	PD 10-AUG-1999		
	PF 12-DEC-1997 JP 1997342899		
	PR		
	PI EJIRO WYANABE, KENJI OEDA		
	PC C12N15/09, C07K14/415, C07K16/16, C12N1/21, C12N5/10, C12N9/88, PC		
	C12O1/527,		
	PC C12O1/68, G01N33/573//C12N15/09, C12R1/91), (C12N1/21, C12R1/19)		
	PC C12N5/10, C12R1/91), (C12N9/88, C12R1/91), (C12N9/88, C12R1/19),		
	PC C12N5/00, (C12N5/00, C12R1/91), (C12N5/00, C12R1/91) CC		

FEATURES		source		Location/Qualifiers	
CC	Topology: Linear	EH	key	FT	CDS
Query Match	41.1%	Score 966:	DB 6:	Length 2497:	
Best Local Similarity	66.0%	Pred. No. 4.3e-277:			
Matches 1507:	Conservative	0:	Mismatches 740:	Indels 37:	Gaps
<div> <div>Strandedness: Single:</div> <div>Location/Qualifiers</div> <div>62..2407.</div> <div>1..2497</div> <div>/organism="Glycine max"</div> <div>/db_xref="taxon:3847"</div> </div>					
BASE COUNT	650 a	521 c	673 g	653 t	
ORIGIN					
57	TCACATGTCGTCACCGTTGGCAATCGACGACGATCGATTTACCTGTGAGACGCTATCTCGTT	116			
Db	118 TAAATTTGCCCTTTGTCCATTAACCCCTAGAGATCAATTTTCTTCGCCAACGGCCACCTTTT	177			
Qy	117 TCTGTCCGATGTTTCTTGAGACATGTTGTCCTTCTCTTCCTGCTAC-----AC	164			
Db	178 TCTCAGCGAAGTTCCCGAAACATATATAGTACCCCTTACCCATCGACGCCAAGATAG	237			
Qy	165 TTTCGATAGACAGTCCCGGTTTCGGTTGCTGTGCTTTGTTGATTCGACGCGTCGGAACC	224			
Db	238 TAAAGAACACGAGGACGACGACGCTGCTAGGTTGCTTCGCGGCTTCACGCGACGAGCC	297			
Qy	225 TCGATGCCGACATGTTGTTTTCGATTGGGAAAGCTGGAAGATATTCCGTTTATGATATTTT	284			
Db	298 CAGAACCCGACACGCTGGCTTCCCTGGGGAAGCTCAGAGGAATTAATTAATTCATGACATATT	357			
Qy	285 CAGGTTTAAAGTTGGTGAGCTACACACTGGGTTGGTCGAAATGTGGGATCTTGAATC	344			
Db	358 CCGGTTTAAAGTTGGTGAGCACCTACCTAGGCTGAGTACCAAGCACGACGAACCTGGAGCA	417			
Qy	345 GGAGACTCAGATTTGTGATCCTTGAGAACTGAC---ATTGGTGCACCGTATGTTTCT	401			
Db	418 CGAGACACGATGATGCTTCTCGACAAAAGACACGCTCGGACGCGCCCTTGTGTTGAT	477			
Qy	402 TCTTCCGATCGTTGAGGAGCAGCTTCGGAACCTCGATTGACCGCGGGATGATGATCTTGT	461			
Db	478 TCTCCCGATCCTTCACAGCCTCGTTCCGAGCCTTCCTCGCAACCCGGTTTGGATGATTAACGT	537			
Qy	462 CGATGTTGTGTGAGAGAGTGTTCGTGCAAGTGTGATGATGATCGTTCGAAATGATT	521			
Db	538 GGACGTTTGCATGAGAGAGCGGGGTGCACAGCTGTCTGTGGCTCCAGCTTGGGAGCTGCTT	597			
Qy	522 GTATCTTACGCTGTGATGATCCGTTTGCACTTGTTAAGAGGCGATGAAGTCTGTAG	581			
Db	598 ATACGTCACGTTGGGCATGACCCGATATGATGTTGCTTAAAGAAACAATAAAGTCTGTAG	657			
Qy	582 GACCCATCTTGAACCTTTTCGCTTGTGGAGAGAGAACTCCACACGATATCTGTGACAA	641			
Db	658 GATGCAATTTGGGAGCGTTTCAAGCTTCTCGAGAGAAACCGCGCGATATCATTAACAA	717			
Qy	642 ATTGCGTTTGGTGCACGTTGGAGCGCGTTTACCTTAACGATTCATCCACAGGCGCTAATAGA	701			
Db	718 GTTTTGTGTTGTACATGSGAGCGCTTTTACTTGAAGTGTGATCCTTCAGTGTGTGGGA	777			
Qy	702 AGGCGTAGGCAATCTGCTGCAGCGCGGTTGTCTCCCGGTTTATGTCCTAATCGACGATGG	761			
Db	778 AGGGGTGAAGAGTTGTGTGGAGGAGGAGGCGCCCTCCAGGAGATGTCTAATCGACAGCG	837			
Qy	762 TTGGCAATCCATCGGACACGATTTGGATCCCATCNC---CAAGAAGATGATGACCAAC	818			
Db	838 GTGGCAGGCAATTTGTGCACGAGAGGAGACCCCATTAACGACGACAGAGGATGTGAAGGAAC	897			
Qy	819 CGTGCGCGCGGAGCAAAATCCCTGCGCTTTTGTGAATTCACAGAAATTTACAAATTCG	878			
Db	898 CTCGCGAGGAGGAGCAAAATCCCATGAGTGTGGTGAAGTGGAGGAAATTTACAAAGTTCA	957			
Qy	879 TGACTATGCTCAATCCCAAGGCCACCGGCCCCCGAGCGGCCAGAAAGGGAATGAAGCGCTT	938			

Db 958 ACAGTATTGATGGAAAGATTCTG-----AGAAAGGTAATGGGTGCTT 1002
 QY 939 TATAGATGAACCTCAAGAGAGATTAAAGACTGTGAGACATGTTATGTTGGCATCTTT 998
 Db 1003 TGTATAGGACTTGAAGGAACAGTTTGGAGCCGTGGAGACAGTGTATGTGTGGACCGCT 1062
 QY 999 GTGTGATATGGGGTGGCCCTTGCCTCCGACAGTGTCCCTGGCTGAGCAGCTGTGAT 1058
 Db 1063 TTGTGGGTATTTGGGGTGGGTGAGACCAGAGTTCGCGGATGCCAGGCTAAGGTGT 1122
 QY 1059 TCAGCAGAGCTTTACACAGAGCTGCAGATGAGAGATTTGGCGGTGATTAAGT 1118
 Db 1123 CACTCCGAAAGCTCCAAATGACTAAATTAATGAAATGAAAGATTAGCGGTGATTAAGT 1182
 QY 1119 TGTCTTTCATTAAGTGGGGTGTCCCGCGGAGAGAGCTGAGAGATGATGACAGAGACT 1178
 Db 1183 CGTCAGTTAAAGGAGTTGGAGCTGTGGCACACACCTGGCTCACCTTTTGTAGCAGGGCT 1242
 QY 1179 TCATGCTCATTTGGAAAAAGTTGGATCGACGCTGTTAAGATTGACGTTATCCACTATT 1238
 Db 1243 CCACCTCCGTTTGGAAATCTGCGGGTATTGACGGTGTAAAGTTAGAGTTATACACTTGGT 1302
 QY 1239 GGAGATGTTGTGAGAGACTATGAGAGAGAGATTTGGCAAGGCAATATTACAAAGC 1298
 Db 1303 CGAGATGCTATCGAGGAATACGGTGGCCGTGTGAGCTAGCCAAAGCTTATTACAAAGC 1362
 QY 1299 AATGACCAATCAATTAATTAATTAATTAAGGAATGAGATGATGCAATGATGAGACA 1358
 Db 1363 GGTCACTGCTGTGGTGAAGAGAGATTTCAAAAGCATGGGGCTATTGCGAGCATGAGACA 1422
 QY 1359 TTGTAAAGCACTTCATCTCTTCTGACACGGAAGTATCTCTTGTGTGTGTGTGATGA 1418
 Db 1423 TTGTAAATGACTTCTTCTCTGTTGACCGAAGCATAGCCTTGGCGGTAGAGATGA 1482
 QY 1419 CTTTGTGACAGGACCCCTGTTGATTCGCAAGAGTACGTTTGGCTCCAAAGATGCTA 1478
 Db 1483 TTTTGTGACAGTACCTCTGTGAGATCCAAATGACAGTATGCTCCAAAGGTCTCA 1542
 QY 1479 CATGTTTCATTTGTGCAACGACAGCTTGTGATGGGAACTTCATCCACTGACTGGGA 1538
 Db 1543 CATGGTACACTGTGCTTACAAACAGCTTGTGATGGGAAATTTTATTCAGCGGATTTGGGA 1602
 QY 1539 TATGTTCAATCCACCACCCTTGTGCGGCTTGCATGCTGCTCTGAGCCATCTCTG 1598
 Db 1603 CATGTTCCAGTCCACCTGCTGACCCATTCATGTC-GCCTCTAGGGCCATCTCTG 1661
 QY 1599 TGGCCGATCTATAGTATGATCTGTGGAAAGCATTAATGATCTCTGAAAAACT 1658
 Db 1662 TGGACCAAGTTTACGTTAGTATGATTTGTGAAAGACACAACTTCAAGTTGCTCAAGACCT 1721
 QY 1659 AGTGTCTCTGATGATGATGATCTCTGAGTGAAGTATGATGATGATGATGATGATG 1718
 Db 1722 CGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1781
 QY 1719 TTTTGTGAGACCTTTCATTAATGAGAACTATGCTTAAGATTGGAATCTCAACA 1778
 Db 1782 TTTGTTTGAAGACCTTTCATGATGAGAAAGCAATGCTCAAAATTTGGAATCTCAACA 1841
 QY 1779 GTTCACTGAGATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1838
 Db 1842 AATATAGGTTGTTGGGTCTATTTAATGATGATGATGATGATGATGATGATGATGATG 1901
 QY 1839 CCGCAACCAATGCTTTTCAACAATACTCAAAAGGATGATGATGATGATGATGATGATG 1898
 Db 1902 GAGAAACAGAGTGGCTCTGAATTTTCAACAATGATGATGATGATGATGATGATGATG 1961
 QY 1899 CATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1958
 Db 1962 CATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2021
 QY 1959 TTACCTTATCAAGCAAAACTTATCTCTCAAGCCCTCTCAAGATCTTGACATAGC 2018
 Db 2022 AATATTTGTTCAAGGACCAAACTAAAGCTCATGAAGGATCATGAGAAATTTGGAAGTTTC 2081

QY 2019 TCTTGACCATTCGAATTCGAGCTCATCTGTTTCACAGTGACCAACTCATCCAAAC 2078
 Db 2082 ACTTGAGCCATTTACTTTGAGCTATTTAGACAGTGTCTCCAGTGATTTGTGTCAAAAA 2141
 QY 2079 TTCTCTACACTTTGCCCAATTTGGCTGGTGAACATGCTTAAACATTAATGAGACATCCA 2138
 Db 2142 GTTAATTCATTTGCTCCCAATTTGATTAAGAACATGCTTAACATGCTGTGCTCATTTCA 2201
 QY 2139 ATCTGAGCATATGAGCATATGACCTAAGCTCATGCGATGATGATGATGATGATGATGATG 2198
 Db 2202 GTTCATGAGATTTGACACACCATATGATGATGATGATGATGATGATGATGATGATG 2261
 QY 2199 GATGCGGATTTTGCATTCGAAGAAACCAAGGCTGTGCTATTAATGATGAGGAGATTTGG 2258
 Db 2262 GATGAAGGTGTTTGATCAAGAGAACACAGTTAGTTGCAAACTAGATGGGTACTTGTAAA 2321
 QY 2259 GTTCAAGTATGATCAGACCAAAATGCTGTGTTCAAGTCCCATTTGCAATTTATCTTC 2318
 Db 2322 AATTGATTTATG--AGGATAAATGCTGAGAGTGAAGTTCCTTGCGCTAGTCTTCAAA 2378
 QY 2319 ATCG 2322
 Db 2379 ATTG 2382

RESULT 9
 PSAA26475 2671 bp mRNA linear PLN 15-JAN-2002
 LOCUS
 DEFINITION Pismum sativum mRNA for raffinose synthase (rfs gene).
 ACCESSION AJ426475 GI:18181864
 VERSION AJ426475.2
 KEYWORDS raffinose synthase; rfs gene.
 SOURCE
 ORGANISM Pismum sativum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae;
 Pismum.

REFERENCE
 AUTHORS Peterbauer, T., Mach, L., Mucha, J. and Richter, A.
 TITLE Molecular characterization of raffinose synthase from pea (*Pismum sativum* L.) seeds
 JOURNAL unpublished
 REFERENCE
 AUTHORS Peterbauer, T.
 TITLE Direct Submission
 JOURNAL Submitted (02-JAN-2002) Peterbauer T., Institute of Ecology, University of Vienna, Althanstrasse 14, Vienna, A-1090, AUSTRIA
 REMARK
 REFERENCE
 AUTHORS Peterbauer, T.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2002) Peterbauer T., Institute of Ecology, University of Vienna, Althanstrasse 14, Vienna, A-1090, AUSTRIA
 COMMENT On Jan 16, 2002 this sequence version replaced gi:18071124.
 FEATURES
 source

1..2671
 /organism="Pismum sativum"
 /cultivar="Wunder von Kelvedon"
 /db_xref="taxon:3888"
 /tissue_type="seeds"
 /country="Austria"
 148..2544
 /gene="rfs"
 148..2544
 /gene="rfs"
 /EC_number="2.4.1.82"
 /function="galactinol:sucrose galactosyltransferase"
 /codon_start=1
 /evidence="experimental"
 /product="raffinose synthase"
 /protein_id="CAD20127.2"

CDS
 gene
 CDS

Query Match	Similarity	Score	DB	Length
Best Local	66.9%	Pred. No. 2.9e-274		
Matches 1431	Conservative	0	Mismatches 684	Indels 24; Gaps 4
BASE COUNT	752 a	557 c	620 g	742 t
ORIGIN				
Query Match	40.7%	Score 956.6	DB 8	Length 2671
Best Local	66.9%	Pred. No. 2.9e-274		
Matches 1431	Conservative	0	Mismatches 684	Indels 24; Gaps 4
193	GCTTCCTGTTGATTTGATTCGACCGCGTCGCAACTGATAGCCGACATGTTTTCGATTGGG	252		
Db	GGTTGTTTCGTTGGGTTTCAACACACGGAAGCCAAAGGACACAGTGGTTCCACATCGGC	456		
253	AAGCTGAAGGATATTCCGGTTTATGATATTTTCAGGTTTAAGGTTTGGTGGACTACAC	312		
Db	AAACTTAAGGATTAATTAATTCACGACGATATTCCGGTTTAAGGTTTGGTGGACACTAC	516		
313	TGGGTTGGTCGAATGTTGGGGATCTTGATGTCGAGACTCAGATGTTGATCTCTTGGAAG	372		
Db	TGGGTCGGAACAACGACACGACACTACAAACAGAAACCAATGTTAATCTCGACAAA	576		
373	TCAGATTCC---TGTCGACCGGATGTTTTCCTTTCGATCGTTGAGGACCGTTCCGA	429		
Db	AACATCTCCCTCGGACGACACTTCCCTCCCTCCATCCCTGMAAATCCCTTCCGA	636		
430	ACCTGGATTTCACCGCGGGGATGATGACTTGTGCAATGTTGTTGGTGGAGTGGTTGGTG	489		
Db	ACCTGACTTCCACCGGCGCTCAACGATTAACGTCGACATATCCGTCGAGACGGTTTACAA	696		
490	AAAGTTGTTGATGATCGTTCCGAAGTATGTTGATCTTCATGCTGGTATGATCCGTTT	549		
Db	CATGTCACCGGTTCAACCTTCAAGCGTCTCTTTACCTCCATCTGACGAACGACCATAC	756		
550	GCACTTGTTTAAAGAGCGGATGAAGATCGTGAGGACCCATCTTGGAATCTTTCGCTTGTG	609		
Db	CGTTTAGTGAAGGAAGCAGTGAAGTAAATCCAAACCAAGTTAGGACATTTCAAGACTCA	816		
610	GAGGGAAGACATCCACACAGTATCGTGGCAAAATTCGGTTGGTGGACGAGGACGGCTTT	669		
Db	GAGGAAAAAACCCACCTAGTATCATCGGAATAATTCGGTTGGTGGACGAGGATGCGTTT	876		
670	TACCTAACGCTTCATCCACAGGCGGTAATAGAAAGCGGTAGGACATCTCTCCGACGCGGT	729		
Db	TATTTGAAGCTTCATCCAAAGAGTGTATGGGAAGGTTGAAGCTCTCACAGCGGTGGT	936		
730	TGTCCTCCCGGTTTAACTCTAAATCGACAGTGTGGCAATCCATCGGACGATTCGGAT	789		
Db	TGTCCTCCCGGTTTGGTCAATTCGACGAGCGGTGGCATTCATTTCTCATGATGATGAT	996		
790	CCCATTCACCAAGAA-----GGAATGAACCAACCGTGGCGGCGGAGCAATGCGCTCG	843		
Db	GATCTCTTACAGACGAGATGGAATGAACCGAACTCAGCTGGGAGAACAAATGGCATATC	1056		
844	CGCTCTTTGAATTCACAGAGAATTCACAATTCGCGTACGTACGTCAATCCCAAGGCGCAC	903		
Db	AGCCTTATTAATATGAGAGAATTTTAAATTTAGAGAATACGAAAT-----1104			
904	GGCCCCGAGCGCGCCAGAAAGGAGATGAAGCGGTTTATGATGAACTCAAGAGAGATT	963		

D	b	1105	GGCTATTAATGGAGGAGAAAGGTTTGTTGGTTTGTGTGATGGGATTTTGAGAGACATTTT	1164
O	y	964	AAGACCTGTGGAGACATGTTTATGTTTGGCATGCTTTTGTTGTGGATTAATTGGGTGGCCCTTGC	1023
D	b	1165	AGGAGTTGTGAGAGTGTTTATGTCTGGCATGCCCTTTGGGGATTTGGGGTGGGTTTGA	1224
O	y	1024	CCGAGGTGCCCTGGCTTGCCTTGAGGACGTGTGATTTACGCCAGTGCCTTTACACAGGCTG	1083
D	b	1225	CTTAAGTGTGTGGAAATGCGCGAAGCTAAAGTGTGTGTTCCGAACCTGTCTCCGGGGTG	1284
O	y	1084	CAGATGACGATGGAGGATTTGGGGGTGGATAGATTTGTTCTTCATTAAGGTGGGCTGTG	1143
D	b	1285	AAGATGACGATGGAGGATTTGAGGGTGGATTAAGATTTGTTGAGAAATGGTGTGGGCTAGTG	1344
O	y	1144	CCGCGGAGAGAGCTGTGAGAGATGTACGAAGACCTCATGCTCATTTGGGAAAAAGTTGGG	1203
D	b	1345	CTCTCAAAATTTGGCGCGAGAGATGTTTGATGGGATTTCACTCTCATTTTGSAGTGGCGGA	1404
O	y	1204	ATGAGCGGTGTTAAGATTGACGTTATCCACTATTGGAAGTGTGTGTGAAGACTATGSA	1263
D	b	1405	ATTGACCGTGTAAAGTTGACGTTATCCATTTGCTTGAGTTACTATACAGAGAAATATGTT	1464
O	y	1264	GGGAGATGCAATTTGGCAAGGCATATTTACAAAGCAATGACCAATTCATTAATTAACAT	1322
D	b	1465	GGAGAGTGTGAGTGAAGAAAGCTTATTTACAAAGCACTAACTCACTCACTGAAACAAAT	1524
O	y	1324	TTTAAAGAAATGGAGTCAATTGGCAAGTATGGACATTTTGAACGACTCATTTGTCCTTGGC	1383
D	b	1525	TTCAAGGCAATGTGTATTTGCAAGCATGGACATTTGCATATGACTTCTTCTCTCGGC	1588
O	y	1384	ACGGAAGCTATCTCTCTTGGTGTGTGTGATGACTTTTGGTGACGAGACCCCTGTGT	1443
D	b	1585	ACGGAAGCATATCTCTCGCGCGCGTGGAGATGATTTTGGTGCTGTGACCCCTGTGT	1644
O	y	1444	GATCCAAACGCTACGTTTGGCTCCAAAGATGTACATGCTTCATTTGTGCCAAGCAGAC	1503
D	b	1645	GATCCAAATGTTATATTTGGCTCCAAAGTGTGCATAGTGTACATTTGTCCCTACAAAGT	1704
O	y	1504	TTTGGATGGGAACTTCATCCACCAGTGTGGATATGTGCCAATCCACCACCCCTGTG	1563
D	b	1705	TTATGTGATGGAAATTTTCATTCATCCAGATTTGGACATGTTTCAGTCCACTCATCTGT	1764
O	y	1564	GCCGCTTTCATGCTGCTCTCGAAGCCATCTGTGGGCCGATCTAATATGATTTCT	1623
D	b	1765	GCCGATTTTCACGCGCGCTCAAGAGCCATCTGTGGACACAGTTTACGTTAGTGATTC	1824
O	y	1624	GTGGGAAGCACTAATCTTGATCTTCTGAAAAAACTAGTCTTCTGTATGGATGCATCTT	1683
D	b	1825	GTTGTATATCAATTTTCAAGTGTCTCAAAATCTTTTGTGGCCGACGTTCTATCTTG	1884
O	y	1684	CGAAGTGTCTATGTGACATCCGACCTCGGATTTGTTTGGTAAGACCCCTTGGATAT	1743
D	b	1885	CGTTGTCAACATTAGCACATCTCCACGAGACTGCTTATTTTGAAMACCCCTTGGCATAT	1944
O	y	1744	GGAGAACTATGCTTAAAGATTTTGGAATCTCAACAGTCACTGGAGTGAATGGTGATCT	1803
D	b	1945	GGCAAAACAATGCTCAAAATTTTGGAAATCTCAACAATATGTGTGTGTTTGGGTGATTC	2004
O	y	1804	AACGTCCAAAGAGAGAGATGTGTCTGTAGACACGCCCAACCAATGCTTTTTCACAATTC	1863
D	b	2005	AACGTCCAAAGTGTGTGGTGTGTCTCGAAGACGCCGAACAGAGCCCTCTGTGATTT	2066
O	y	1864	TCAAAAGSAGTGAATCCAAACATCAACCCAAAAGACATGAATGGCAAGTGGAGAAAC	1922
D	b	2065	TCACACGCGGTGAATGTTATGCAAGTCTGGAAGATTTGAATGGTGTGCATGTAAACAC	2124
O	y	1924	CTTATCTTATTAAGAGCGTTAAAACTTTGGCTTTTACCCTGTATCAACCAAAAAACTT	1983
D	b	2125	CCAATGAGACATCAAAAGGTGTGATGTTTGTGCTGTGATTTTTCAGAGAAAGAAACAG	2184
O	y	1984	ATCTCTCTCAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCAATTCGAATTCGAGCTC	2043
D	b	2185	AGTCTCATAGAGTCTTCGATAGATTGGAGATTTTCCCTCGAGCACTTATAGTTTATAGCTA	2244

QY	1390	GCTATCTCTTGGCGGTTGGTGGATACATTGGTGGTACAGGACCCCTGGTGATCA	1449
Db	1547	GCATATCCCTCCGGCCGGGTGGAGATATTTTGGTCTCTGATTCATCTGGTGATCCA <td>1606</td>	1606
QY	1450	AACGGTACGTTTGGCTCCAAAGATGTCCATGTGTTCACTTGTGCCAACGACGCTTGG	1509
Db	1607	AATGGTACATATATGGCTCCAAAGGTTGTACATATGGACATTTGGCTTACAAAGTTATAG	1666
QY	1510	ATGGGGAACCTTCATCCACCCTGACTGGGATATGTTCCATATCCACCACCCTTGTGCCCC	1569
Db	1667	ATGGGAATTTTCATTCACCCAGATTTGGGACATGTTTCAGTCCACATCATCTTGTGCTGAA	1726
QY	1570	TTTCATGCTGCTCTCCGAGCCATCTCTGGTGGCCGATCTATGTTAGTATTCCTTGGGA	1629
Db	1727	TTTTCATGCGCCCTCAGAGGCCATATCCGGCGGACCAATTTATGTTAGATTTGTGTGT	1786
QY	1630	AAGCATACCTTTGATCTTCTGAAAAAACTAGTCTTCTCTGATGATGATCCCTTGAAGT	1689
Db	1787	AATCAATTTCAAGTTGCTCAAAATCTCTGTGTTGGCCGATGGTTCTATCTTGGCTGT	1846
QY	1690	GAGTACTATGCACTCCGACATCCGATTTGTTTGTGGAAGACCTTTTGATATATGAGAA	1749
Db	1847	CAACATTTACGAGACCTCCATCAAGAAATTTGTTTGAAGACCTTTTCATATATGGCAA	1906
QY	1750	ACTATGCTTAAGATTTTGGAAATCTCAACAAGTTTCATCGAGATGATGTTGATCAACTGC	1809
Db	1907	ACATATGCTGAAATTTTGGAAATCTCAACAATATATACAGGTGTTTGGGTCTTTTCAACTGC	1966
QY	1810	CAAGAGAGAGATGTGTGCTGAGACACGCGCAACCAATGCTTTTTCATATACTCAAA	1869
Db	1967	CAAGTGTGGTGGTGTGCTGAGACGCGGCAACCAAGAGTATCTGAATTTTCAAGC	2026
QY	1870	CGAGTACATCCAAACCTAACCCAAAGCATAGATGAGACAGTGGAGAAAACCTATC	1929
Db	2037	GCAGTACATCTTAAAGCAATGCCCAACATGTAATGGTGCATATGGGAAAACTCAATG	2086
QY	1930	TCTATTGAAGCGTTAAACCTTTGCGCTTAACTCTATCAAGCCAAAACTTATCTC	1989
Db	2087	AGCACCAAAAGGTGTGATTTTTCCTGTGTATTTTTCAGAGAAAGAAATGAGGCTC	2146
QY	1990	TCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCCATTCGAATTCGAGCTCATCT	2049
Db	2147	ATGAGTGTCTGATAGATTGAAGTTTCGCTTGAGCCTTGAATTTGAGTAAATGACA	2206
QY	2050	GTTTACCAGTGAACCAACTCATCCAACTCTCTACACTTTGCCCAATTTGGGCTGGT	2109
Db	2207	GTTGTTCCAGTGAAGTGTGTTGAAAAAGGTTTAAACAGTTTGGACCGATGGGTTAGT	2266
QY	2110	AACATGCTTAACTAGTGAAGCCATCCAACTGTGGACTATGACAGTACCTAGCTCA	2169
Db	2267	AACATGCTGAACCTGTGGTGTGCTATTAAGTCTCTGGAGTTTGAATGATATCAAGTTTG	2326
QY	2170	GTCGAGATTGTGTCAAAAGGTGTGTAGATGCGAGTATTTGATCGAAAAAACCAAG	2229
Db	2337	GTCAGATTTGGGCTGAGAGGTTCGCGGAGATGAGACGTTGCTGTGAGAAACCGGTT	2386
QY	2230	GCTTGTCTGATTTGATGGGAGAGATTTTGGGTTCAAGTATGATCAGACCAATTTGTTGGT	2289
Db	2387	TGCTGTCAAAATTTGATGGGCTTAAAGTGAATTTCTTTATG--AGCACAAATATGGCAAGA	2443
QY	2290	GTTCAAGTGCATGGCCAAATGATCTTCAATCG	2322
Db	2444	GTTCAATTTCTGTGGCTAGTTCTTCAACATTG	2476

SOURCE
ORGANISM
Arabidopsis thaliana.
FLI.CDNA.

REFERENCE
AUTHORS
Sakurai,T., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Saitou,M., Seki,M., Shinozaki,K., Becker,J., Theologis,A. and Davis,R.W.

TITLE
JOURNAL
Submitted (27-FEB-2002) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

COMMENT
comment 'e-mail for correspondence: arab@sequence.stanford.edu

FEATURES
source
Riken Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL.CDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J., Saitou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PEGC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Becker,J., Theologis,A. and Davis,R.W.

Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

location/Qualifiers
1..2395
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="5"
/clone="U17209"
/note="This clone is in pENTR/SD-dtpro. This is a cloned PCR product using RIKEN clone RAFL06-10-P15 (AY062781) as a template
ecotype: Columbia"
1..2395
/gene="MP012.13"
1..2352
/gene="MP012.13"
/codon_start=1
/product="raffinose synthase protein"
/protein_id="AA10207.1"
/db_xref="GI:20148633"
/translation="MASPCLTKSDSGINGVDFTFKFRLDSTLLANGOVYITDVPAVNWTTTHWGSNPLVDKGVLPILVASAGSEFGPLNDGPGSHYASLGKINIFMSFRKWMMADTTHWGSNGLDIENETPIITILDOSGSPRGSGGRVYLILLPLESRFSRPSQGEDDDVAACVGESSTEVTGESEFRIVTVHAGDDFFKLVAQAKVIRAHMNTFLIEKSPPGVIDKFGMGDMDFATLYLVNPDPGVKRVKCVLDGCPGLVLIDGWOSIGDSGIDV EGNMTIVAGEDMDPRALKLFBNHKAFFKDYSPKDNDGMKNAFDLDESFVDY IYVVWHAITCGYWGCLRPAPALPVSTIIRELPELTLMEDLAWDK ILFIIGIASPDLLAKKEFYVALGISHLONAGIDGVAVDTIHILELKOKYRGVRLAKAFAKALTSSVNKFPCNGN GVIASMBCNDMEFMELTGIAISIGVGDGPFCTDGSDGPNCFMILGGCHWVCANSIM LMGNTIPDMDMFQSTHPCAERHASRAISGPTIISGCVKHNDLKRVLPMGSTILL RCETALPTDRDLFEPLHDHCKTLKMLNKRYGVGFACFCQGCGMCREFTRNNQCS ECUTVLAATTSPKDEVNNSGSPISLIANVEEFALELSQSKLLSLGINDLELETLEPP FKEFLITVPVVTIGENVRFAPICGLVMNLNLSGAIRSLVYNDESVEGACGEFRVY ASKPVSCLIDGEEVVERGVEDSNVMOVVMSGPDGLSIOTLF"

BASE COUNT
ORIGIN
566 a 532 c 602 g 695 t

Query Match 40.4% Score 950.6; DB 8; Length 2395;
Best Local Similarity 67.4%; Pred. NO. 1.7e-272;
Matches 1422; Conservative 0; Mismatches 644; Indels 45; Gaps 4;

RESULT 12
AY062781 2718 bp mRNA linear PLN 25-NOV-2001
LOCUS Arabidopsis thaliana raffinose synthase protein (MPO12.13) mRNA,
DEFINITION complete cds.
ACCESSION AY062781.1 GI:17065409
VERSION FL1.CDNA.
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots: Rosidae: eurosids II: Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 2718)
AUTHORS Southwick, A., Karlin-Neumann, G., Nguyen, M., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H., Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A., Kawai, J., Kim, C., Lin, D., Liu, S.X., Narusaka, M., Pham, P.K., Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K., Hayashizaki, Y., and Shinozaki, K.
TITLE Submitted (14-NOV-2001) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT e-mail for correspondence: arabesequence.stanford.edu
RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: "RIKEN Arabidopsis Full-length cDNA"): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.
The Sak, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen, M., Southwick, A., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K., Kim, C., Lin, D., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.
Southwick, A., (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.
FEATURES
source Location/Qualifiers
1..2718
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="v"
/clone="RAFL06-10-P15"
/note="this clone is in bluescript
ecotype: Columbia"
1..2718
/gene="MPO12.13"
132..2483
/gene="MPO12.13"
/codon_start=1
/product="raffinose synthase protein"
/protein_id="AAL32859.1"
/db_xref="GI:17065410"
translation="MASPLTRKSDSGINGVDFTEKRLFEDSTLLANGVLTVDVYV
TLTSPLYVDKGVPLDYASGIFGFIENLDGPKSHVASTGLKIRMSIFRQVW
TTHWGSGRDIENETQIILIDSGSDSGPSGSRPVLLPLLEGFSRSPQED
DDVAVCESGTEVGESEFROIIVYHAGDDPKLAKDMKIVRMATFKLEESPP
GIVDFGCTMDAFLYLVNPDGVHKGVCGLVDGCPPLVDLDMOSIGHSDSIDV
EGMTVAGDEMPRLKFEENHKFKDYVSPKDDVMKMFVRDLKDFEFTVDIYV
MHALGCVGGLRPPRALPSPITIRPELSPGLKTLMEPLAVDKITTEIGTASPLAK
EYEBGLSHLONAGIDGVKVDYIHLLELCKYGGRVDLAKATFRALTSVKNKFN
GVASMEHNDPMELGTALISLGRVDFWCTDPSDNGTFWLOGCHMVCANSLW
MGNFIQDPMDFQSTHPCAEFRAISGEPYISDCVGHFDPDLRLVLPNSIL
RCVYALPTDRLEFDPLHDKTLMKINLNKYTVIGALVNCOGGCMRERTRNCF
ECVNTLATPTSKDEVNMSGSPISIANVEEPALFLSOKKLLSGLDLLETLEPF
KPELTIVSVYTTIEGNSVPAPICGVNLTNSGATRSIVYNDSEVGVFCAGERYV
ASKRVSCLLIGEVYERFEDSMVNVQPMSPDGLSTIYLF"

BASE COUNT 658 a 581 c 643 g 836 t

ORIGIN

Query Match 40.4%; Score 950.6; DB 8; Length 2718;
Best Local Similarity 67.4%; Pred. No. 1.8e-272;
Matches 1422; Conservative 0; Mismatches 644; Indels 45; Gaps 4;

QY 220 GAACCTGATAGCCGACATGTTGTTGATTTGGGAAGCTGAAGATATTGCTTTATGAGT 279
DB 354 GAGCCCAAAAGCCACACAGTGCATTCATCGGAAAACACAAACATTCGTTTCATGAGC 413
QY 280 ATTTTCAGGTTTAAGCTTTGGTGGACTACACACTGGGTTGGTGAATGCTGGCATCTT 339
DB 414 ATATTCGTTTCAAGTTGTTGGAGTACTCATTTGGTGGATCAACAGGACGTGACATC 473
QY 340 GAATCGAGACATCAGATTGTGATCTCTTGAGAA-----GTCA 375
DB 474 GAGAACGACACTCAATCATCTTCTTGATCAATCCGGGTCCGATTTCTGAGACCCGATCC 533
QY 376 GATTCTGTGCGACCGTATGTTTCTCTTCCGATTCGTGAGGAGACCTTCCGAACCTCG 435
DB 534 GGGTCGGGCGTCTTATGTTCTCTTCTGCTACTCTTCTTGAAGGCGCTTCCGTTTCATCA 593
QY 436 ATTCAGCCTCGGAGATGATGACTTTTCCATGTTTGTCTGAGAGTGTTCGTCAAGATT 495
DB 594 TTCCATCCGGAGAACAGACGATGACGTGGCGGTCTGTCTGTAATCCGGGTGACCGAAGTA 653
QY 496 GTTGATGATCGTTCGCAAGATGTTGTAATCTTCATGCTGGTGAATGATTCGTTGACATT 555
DB 654 ACCGGGTGCGAATTTGTCAGATTTGTATGTTCATGCGGGGATGACCCGTTCAAGCTC 713
QY 556 GTTAAAGAGCGCATAGATTCGTGAGAACCCATCTTGGAACTTTTGGTTTGGAGAG 615
DB 714 GTGAAGAGCCCATGAAAGTATTAGGCTTCAATATGATTAAGCTTCAAGCTTCTGAAG 773
QY 616 AAGACTCCACGAGTATGCTGACAAATTCGTTGGTGCACGTGAGAGCGCTTTTACCTA 675
DB 774 AATGCGCCCGCGGAATGTCATTAATTCGGGTGTGACATTTGGATGCTTTTACTTG 833
QY 676 ACGGTTCATCCACAGCGGCTATAGAAAGCGGAGCAATCTGAGCGGCTGCTCT 735
DB 834 ACGGTGAATCTGACGAGATTCATAAAGGTGTAAGTCTCTGACGCGTGTGCTCT 893
QY 736 CCGGTTTATGCTTATCAGCATGATGTTGGCAATCCATCGACAGATTCGATCCATC 795
DB 894 CCGGATGCTTCTTATGACAGACGTTGGCAATGATTTGACATGATTCGATGAT 953
QY 796 ACCAAAGAGATGACCAACACGTCGCGGCGAGCAAAATCCCTGCCCTTTTGA 855
DB 954 GATGTCGAAGGATGATATTTACCGTCGCGGCTGGAACAAATGCTTGACGCTTGAAG 1013
QY 856 TTCCAAGAGATTTACAATTCGCTACGTCATCCCAAGGCCACGCGCCCGGAGCC 915
DB 1014 TTTGAAGAGAACCAATTTCAAGACTAGCTTTCTCCAAAG-----ATCAA 1061
QY 916 GGCAGAGAGGATGAGAGCGCTTATAGATGAATCAAGAGAGATTAAAGCTGTGAG 975
DB 1062 AAGCAGCTGCGAATGAACCTTTGTCAGAGATGCAAGATGAATTCACACCGTTGAT 1121
QY 976 CATGTTTATGTTTGGCATCTTTGTGTGATATTGGGGTGGCTTCCCGCAGATGCTT 1035
DB 1122 TGCATCTTACGTTTGGCACCATCTTGGGTTACTGGGGAGCTCCCTCAACCTCGC 1181
QY 1036 GCGTTGCTGAGGACAGTGTGATTACAGCCAGTCTTCAAGAGGCTGAGATGACGATG 1095
DB 1182 GCTTCGCCCGCTGACCATTTATCCGCGCCGAGCTCTCTCAGGACTTAACTAACATG 1241
QY 1096 GAGATTTGGCGGTGATTAAGATTGTTCTTCAATAGCTCGGCTGCTCCCGCGAGAG 1155
DB 1242 GAAGATCTCGCGGTGATTAAGATCATAGAGACCGGATGGATTGGCTCGCGCGACTG 1301
QY 1156 GCTGAGAGATGTAGAGAGACTTCACTGCTATTGGAAAAAGTTGGATTCACAGCTGTT 1215
DB 1302 GCGAAGAGTTTACGAAAGTCTTCACTCTCATCTTCAAAAGCGCGGTATTGACGCGCTT 1361

[illegible]

QY	2296	GTCCATGGCC	2306
	11111111		
Db	2433	GTCCCTTGCTC	2443
RESULT	13		
E24425		1762 bp	DNA
LOCUS	E24425		Linear
DEFINITION	Raffinose synthase gene and use thereof.		PAT 18-JUN-2001
ACCESSION	E24425		
VERSION	E24425.1		
KEYWORDS	JP 199215984-A/3.		
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 1762)		
AUTHORS	Eljiro,W. and Kenji,O.		
TITLE	Raffinose synthase gene and use thereof		
JOURNAL	Patent: JP 199215984-A 3 10-AUG-1999;		
COMMENT	SUMITOMO CHEM CO LTD		
	OS Stachys sieboldii (Chinese artichoke)		
	PN JP 199215984-A/3		
	PD 10-AUG-1999		
	PR 12-DEC-1997 JP 1997342899		
	PI PI		
	EIJIRO WYANABE, KENJI OEDA		
	PC C12N15/09,C07K14/415,C07K16/16,C12N1/21,C12N5/10,C12N9/88, PC		
	C12O1/521,		
	PC C12O1/68,G01N33/573//C12N15/09,C12R1:91),(C12N1/21,C12R1:19),		
	PC (C12N5/10,C12R1:91),(C12N9/88,C12R1:91),(C12N9/88,C12R1:19),		
	PC C12N15/00,		
	PC C12N5/00,(C12N15/00,C12R1:91),(C12N5/00,C12R1:91) CC		
	Strandedness: Single;		
	CC Topology: linear;		
	FH Key	Location/Qualifiers	
	FT peptide	2..1762.	
FEATURES			
source	1..1762		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	395 a	451 c	499 g
ORIGIN		417 t	
Query Match	37.7%	Score 886.8	DB 6; Length 1762;
Best Local Similarity	70.0%	Pred. No. 2e-253;	
Matches 1223;	Conservative	0; Mismatches 517;	Indels 6; Gaps 2;
QY	324	AAATGTTGGGATCTTGGATCGGAGACATGATTTGTATCCCTTGAGAACTGATTTCTGG	383
Db	4	AAACGGGTCGGATCTTGAGCGGGAACATCAATAGTCGTGCGACGAAGTCCGA---CGA	60
QY	384	TGCAACCGATATGTTTCCCTTTCGCATCGATCGTTGAGGAGCCGCTTCGAACTCGATTACCC	443
Db	61	CAGGCCCCACATCGCGCTGCTCCGCTCATGAGAGGGGAGTTTCGGGCTTCGCTTCACCC	120
QY	444	TGGGATGATGATGCTTTCGCATGTTTGTGTGAGAGTGCTTCGTCGGAAGTTGTTGATGC	503
Db	121	CGGTGTGATGATTTTATTCGATATTTTGTCTGCAAGCGGGTCAACCAAGGTCAACGAGTTC	180
QY	504	ATCGTTCGGAAGTATGTTGTATCTTCATGCTGCTGATGATCCGTTTGACACTTGTGTTAAAG	563
Db	181	CTCGTTCGCTGCTTCGCTGCTACATGACAGCGCGGTGATGACCTTTTACCTGCTGTAAGA	240
QY	564	GGCGATGAAAGATCGTAGAGACCCATCTTGGAACCTTTTGCTTTGTTGGAGAGGAAGACTTCC	623
Db	241	CGCCGTGAAGGTGGGGCGCCACACACCTCGGAGAGCTTCAGGCTGCTGGAGGAGAAACTCC	300
QY	624	ACCAAGTATCGTGTGACCAATTCGTTGTGTGACGTGGAGCCGTTTTCCTAACGGTTTCA	683
Db	301	CGCGGGATTCGTCCAGCAATTTTCGGTGTGACAGTGGGATTCGCTTCTATCCATCAAGTCCA	360
QY	684	TCCACAGGCGCTAATAGAGCGGTGAGGCATTCGTGACGAGCGGTTGTCTCCCGGTTT	743

```

Db 361 GCCCAGCGGCTTTATGAGAGCGGCTGACAGGCGCTGTTGACGGCGGATGTCGCCGCGGCT 420
Oy 744 AGTCTAATCGAGATGGTTGGCAATTCATGCGACAGCATTCGATCCCATACCAAGA 803
Db 421 GGTGTGATCGACGAGGGGTGGCACTGTTGTACAGCAACAGCGGCTACACCGA 480
Oy 804 AGCAATGAACCAACCGTCCGCGGAGCAAAATGCCCTCCCTCTTTGAAATTTCAAGA 863
Db 481 GGGGATGGGAGAAACCTCCGCGGAGAGCAAAATGCCCTGCGACAGTTGATCAATTTGAGAGA 540
Oy 864 GAATTTACAATTCGCTGACATGCTCAATCCCAAGGCCACCGGCCCCGAGCCGCCAGAA 923
Db 541 GAATTTACAATTCGCTGACGAGTACGAGAGCCCAATTAACATGGGCGC---GGCGCCCAATAC 597
Oy 924 GGGGATGAAGCGCTTTATAGATGAACCTCAAGAGAGATTATAGATGCGAGACATGTTA 983
Db 598 GGGGATGGGCGCTTTATGCGATGAGAGCAAAATTCAGATGAGTGGAGCTAGCTGTA 657
Oy 984 TGTTCGACATGCTTTGTGTGATATTGGGGTGGCCCTGCGCCGAGTGCCTGCGCTGCC 1043
Db 658 CGTGTGGCATGCGTGTGTGGTATTATGGGGCGGCGCTCAGGCCCAATGTTCCGCGCGCTGCC 717
Oy 1044 TGAGCCAGCTGTGATTCAGACCAGTCTTTACACAGGCGTGCAGATGACAGTAGAGATT 1103
Db 718 CGAGGCTAAGCTCATTTGAGCCCAAACTGACTCTGGGCTTAAGACACCACATGAGAGATT 777
Oy 1104 GGGGATGAATAGTTGTTCTTCATAGAGTCCGGCTGCTCCCGCGGAGAAAGCTGAGAGA 1163
Db 778 GGGTGTGATGATTTGTCACAAATGGCGTGGTGTGGTCCAGCGAGTTGTTGTAACA 837
Oy 1164 GATTCAGCAAGCACTTGTCTCATATTTGGAAAAAGTTGGAGTCGACGCTTAAAGATTGA 1223
Db 838 AATGATGAGAGATTACATTCACATCTGGAATCTGTGGGATGATGAGTCAAAAGTTGA 897
Oy 1224 CGTTTCCACCTATTGAGAGATTTGTGTGAGAGACTAGGAGGAGATGAGATTGGCAAA 1283
Db 898 CGTCATGACATTTGTGGAAGTGTGTGTGAGAGACTAGTGGAGAGTGGACTTTAGCCAA 957
Oy 1284 GGCATATTACAAAGCAATGACCAAAATCAATTAATTAACATTTTAAAGAAATGAGATCAT 1343
Db 958 GGCATTATTACAAAGGCTTATCAAGCTCAAGTTAACAACTTCAACAGGCAAGCGGCTCAT 1017
Oy 1344 TGCAAGTATGAGACATTTGTAACGACTTCAATGTTCTTGGCAGGAACTATCTCTTGG 1403
Db 1018 GCCTGGCGTGGAGCACTGCAATGATGACTTATGTTCTCGGAACCGAGGCCATTACCTGGG 1077
Oy 1404 TCGTGTGTGATGATTTGTGTGACGAGCAACCCCTCTGTGTATCCAAAGCTAGCTTTTG 1463
Db 1078 TCGTGTGGGGATGATTTTGTGTGACATGATCATCTGTGAGATCCCAATGGCAGCTTCTG 1137
Oy 1464 GCTCCAAAGATGTCACATGCTTCAATTTGTGCAAGCAACGACTTGTGGATGGGAGCTTCAT 1523
Db 1138 GTTCGAAAGGATGTCACATGCTGCTACGCGGCTTACAAACATATGATGGATATTTTCAT 1197
Oy 1524 CCACCTGATGAGATATTGTTCCAAATCCACCCACCTTGTGCGGCTTCCATGCTGCTC 1583
Db 1198 CCACCTGATGAGATATTGTTCCAAATCCACCTTGTGCGGCTTCCATGCTGCTC 1257
Oy 1584 TCGACCATCTGTGTGTGCGCGCATCTATGTTAGTATCTGTGGAAAGCAATACCTTTGA 1643
Db 1258 AGCAGCATCTTCGCGGCGGCGCATTTAGCTGATGATCTGTGGAAAGCAACATCTTCCA 1317
Oy 1644 TCTTCGAAAAAACTAGTCTCTGATGATGATGATCCCTGGAAGTGAAGTACTATGACAT 1703
Db 1318 GCTCCTTAGAGACCTGCTTCTTCCGATGATGCTCACTCCGCTTGTGATTTACTAGCGCT 1377
Oy 1704 CCCGACCTGCGATTTGTTTGAAGACCTTTGCAATATGAGAGAACTATGCTTAAGAT 1763
Db 1378 TCCGACTGCGGATGCTCTTTGAAGATCCACTTCAATGAGCAAGACTATGCTCAAAAT 1437
Oy 1764 TTGGAATCTCAACAGATTCATGAGATGATTTGTGATTTCAACTGCAAGAGAGAGATG 1823
Db 1438 TTGGAATTTATACAGATTCACCGAGTGTGCGAACTTTCAACTGCGCAAGTGGCGGCTG 1497

```

```

Oy 1824 GTGTGTGAGACAGCGCGCAACCAATGCTTTTACATATACGCAAAAGAGTGCATCCAA 1883
Db 1498 GAGCGGGAGAGTGGCTCGCAACCAATGCGCTGCCAGATATTCACAGCGGCTCTCTAG 1557
Oy 1884 AACTAACCCAAAAGACATAGATGAGTGCACAGTGGAGAAAAACCTATCTGTTATGAAGCGT 1943
Db 1558 CGCTGGTCGAGATGACATTTGATGATGAGCAAGCAAGAGATGCGATGCAAGTGGAGCGGT 1617
Oy 1944 TAAACCTTTGGCTTTTACCTGATCAAGCCAAAAACTTATCTCTCCAAAGCCTTCA 2003
Db 1618 CAAAACATTCGCGTGTACTACTATTCACGAGAAAGAACTGCTTTTCAACCATCAGA 1677
Oy 2004 AGATCTGACATAGCTCTGACCAATTCGAATTCGAGGCTACAGCTTTTCAACCATGAC 2063
Db 1678 CAAAATCGACATACGCTTGAACCTTCGATTTGAGCTGATGATACCGTTTCTCAGTCA 1737
Oy 2064 CAACCT 2069
Db 1738 AACTCT 1743

```

```

RESULT 14
E36418
LOCUS E36418 1762 bp DNA Linear PAT 18-JUN-2001
DEFINITION Raffinose synthase gene.
ACCESSION E36418
VERSION E36418.1 GI:13019216
KEYWORDS JP 2000014389-A/2.
SOURCE Brassica napus.
ORGANISM Brassica napus.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 1762)
AUTHORS Eljilro,W. and Kenji,O.
TITLE Raffinose synthase gene
JOURNAL Patent: JP 2000014389-A 2 18-JAN-2000;
SUMITOMO CHEM CO LTD
COMMENT OS Brassica napus
PN JP 2000014389-A/2
PD 18-JAN-2000
PF 10-DEC-1998 JP 1998351246
PR
PI EIJIRO WATANABE,KENJI OEDA
PC C12N15/09,A01H5/00,C12N1/21,C12N5/10,C12N9/88,C12Q1/68// PC
(C12N1/21,C12R1:19),(C12N5/10,C12R1:91),(C12N9/88,C12R1:19), PC
(C12N9/88,C12R1:91),C12N15/00,C12N5/00,C12N5/00,C12R1:91) CC
FH key Location/Qualifiers
FT CDS (1)..(1719).

```

```

FEATURES
source Location/Qualifiers
1..1762
/organism="Brassica napus"
/db_xref="taxon:3708"

```

```

BASE COUNT 435 a 384 c 459 g 479 t 5 others
ORIGIN

```

```

Query Match 36.8% Score 866.4; DB 6; Length 1762;
Best Local Similarity 70.6% Pred. No. 2.6e-247;
Matches 1200; Conservative 4; Mismatches 475; Indels 21; Gaps 3;

```

```

Oy 607 TTGGAGGAGAAACATCCACACGATATCGTGACAAATTCGTTGGTGGCGTGGGCGCG 666
Db 1 TTGGAGGAGAAAGAGCGCGCGGGAATCGTCGATTAAGTTGGGTGGTGCACGTGGGATGCG 60
Oy 667 TTTTACCTAACGTTTCATCCACAGGCGCTAATTAAGAAGCGGTGAGCAATCTCTGACGCG 726
Db 61 TTTTATTTACGCTGACACCTGACGAGTTCATTAAGAGGTGTTAAGTGTCTGTCGACGCT 120
Oy 727 GGTGTCTCCCGGTTTATGCTTAATCCAGCATGTTGGCAATTCATGGAACAGATTG 786
Db 121 GGTGTCTCCCGGTTTATGCTTAATCCAGCATGTTGGCAATTCATGGAACAGATTG 180

```



```

/protein_id="CAC38094.1"
/db_xref="GI:13992585"
/translation="MAPLNSTSTNLKTESIFDLSEKFKVKGPFLEPHDVENVSR
SSSICKPSESNAPPSILOKYLAVSHKGGFGFSGHSPDSKSLDPSIF
RPFKMSQWIGKSGSDLOMETONILIVPEPTKSVVILPILEKFRALPPGHDV
KILBSGSTKVESTFNSIATVHFSNEDYDLMEKYSAIRHLSFRLEKTLPNLV
DLFWMCYDAVYLVNPIGIFHGLDSEKGGVEPFIIDGOWSISPDGJHEKEDAK
NLVLGGEGMGRHLRFDECFKFRYESGLGPNSPYDPNNFTLILGJIEHEKLRK
KRELAISKSDLAIESKIKKVVKEIDDLFGEBOSSEKSEKSEYGLAKFTDLR
TKFRGLDYYVHALCGAMGVREPETHLDRKYDCKLSPGIDMEDLAVVEYSKAS
LGLVHPSQANELYDSMSHSLAESCITGVKVRHSLVCEYDGRDLAVVEYEGLT
KSIYKFNNGMISMOKNDPFFELGTRQISMGRGDDPFPDNDPDSGSPWILGVH
MHCYSNLSLMGOMIQPDMDFOSDHVCARHAGSRAICGSP1YVSDMVGSHDPLQV
KLVPDGTIPKCIPIFPLPTKDLKRNPLFDHTTVLKIMFNPKYGVIGAFNCGGMD
PIOMKRGFPECYKPICTVHTEVEEMQKEETSHLGAEEVYVLYNDAEELSLMTK
SEPIOTFOTSPFELSPVPTKLGGLGKFKAPIDGLTNMNSGGTYVLDIEYVNGKIK
VKGGSEFLAISESPKRLQNLNGCEYDFEMLDGLCLVAVPIIEKACVSDMEIFP"
BASE COUNT      809 a      417 c      584 g      917 t
ORIGIN
Query Match      17.3%   Score 406.2: DB 8:   Length 2727:
Best Local Similarity 57.7%: Pred. No. 1.5e-109:
Matches 808: Conservative 0: Mismatches 578: Indels 15: Gaps 4:
QY  920 AGAAGGGATGAAGCGCTTATAGATGAACCAAGAGAGTTTAAAGCTTGAGAGCATG 979
    || || || || || || || || || || || || || || || || || || || || ||
DB  1198 AGTATGATTTGAAGGCTTACACAAAGATTTGAGGACAAAGTTTGAAGTATGATG 1257
QY  980 TTTATGTTTGGCATGCTTTGTGTGATATTGGGGTGGCTTGGCCGCGAGGTCCTGGCT 1039
    || || || || || || || || || || || || || || || || || || || || ||
DB  1238 TTTATGTTTGGCATGCACTATGTGTGCTTGGGGTGGTGTGAGGCGTGAACACACACC 1317
QY  1040 TGCCTGAGGACGCTGATTCAGCAGTCTTTCACACGAGGCTGCAGATGATGAGGAG 1099
    || || || || || || || || || || || || || || || || || || || || ||
DB  1318 TGTATACCAAAATGTGCCCTTGCAAA---CTCTCACTGCTGTGATGAGAACATGAGAG 1374
QY  1100 ATTTGGCGGTGATAGATTTCTTCATTAAGTGCGGCTGCTCCCGCGGAGAAAGCTG 1159
    || || || || || || || || || || || || || || || || || || || || ||
DB  1375 ATCTTGACATGGTGTGATTTCCAAAGCGTCACTTGGGCTAGTTCCTTAGTCAAGCTA 1434
QY  1160 AGCAGATGTACGAAGAGACTTCATGCTCATTTGGAAAAAGTTGGAGTCGAGCGGTAAAGA 1219
    || || || || || || || || || || || || || || || || || || || || ||
DB  1435 ATGAGCTTTATACATCATGATTTCTTATCTTGCTGAAATCCCGTATCACCGGAGTCAAG 1494
QY  1220 TTAGGCTTATCCACCTATTTAGAGATTTGTGTGAACACTATGAGAGGAGAGTGGATTGG 1279
    || || || || || || || || || || || || || || || || || || || || ||
DB  1495 TTGATGCTCATTCATCTCTGTAATATGCTGTGATGAAATGAGAGTATGATGATCTTG 1554
QY  1280 CAAAGCATATTTACAAAGCAATGACCAATCAATTAATTAACATTTTAAAGCAATGAG 1339
    || || || || || || || || || || || || || || || || || || || || ||
DB  1555 CAAAAGTTTACTATGATGAAGATTTGACAAAATTCATTTGTCAAAGATTTTATGCAAAATGAGAA 1614
QY  1340 TCATTGCAAGTATGAGAACATTTGTAACGACTTCATGTTCTTGGACGGAAGCATCTCTC 1399
    || || || || || || || || || || || || || || || || || || || || ||
DB  1615 TGAATGCTAGATGACAACTGTAATGACTTTTCTTCTTGGAACAAAGCAAAATTTCTA 1674
QY  1400 TTGCTGCTGTGGTATGACTTTTGTGTGACAGGACCCCTCTGTGATCCAAAGCGTACGT 1459
    || || || || || || || || || || || || || || || || || || || || ||
DB  1675 TGGGAAGATTTGGGATGATTTTGTGTTCCAAAGATCCAAATGATGATCCAAATGAGGAAGTT 1734
QY  1460 TTTGGCTCCAAAGATGTCATGATTTGATTTGTCACAGACAGCTTGTGATGGGGAAGT 1519
    || || || || || || || || || || || || || || || || || || || || ||
DB  1735 TTTGGTTCGAAGGTGTGACATGATTCATTTGTTCCACAAATAGTTTATGATGGGACAAA 1794
QY  1520 TCATCCACCTGCTGAGGATATGTTTCCATCCACCACCTTGTGGCTGCCTTCATGCTG 1579
    || || || || || || || || || || || || || || || || || || || || ||
DB  1795 TGAATTCAGCTGATTTGGGATGATGTTCCAAATCAAGATCATGTTTGTGCTAAAGTTTCATGCTG 1854
QY  1580 CCTCTGAGCCATCTCTGTGGGCGCCGATCTATGTTAGTATCTGTGGGAAGCATAACT 1639
    || || || || || || || || || || || || || || || || || || || || ||
DB  1855 GTTCAAGAGCTATTTGTGGGCAAAATCTATGTGAGTATATGTTGGCTCTCATGATTT 1914
QY  1640 TTGATCTTCTGAAAAAAGTATGCTCTCTGATGAGATGATCTCTTGAAGTGAAGTACTATG 1699

```

```

|| || || || || || || || || || || || || || || || || || || || ||
DB  1915 TTGATTTGATTAAGAAAGCTTTGATTCGCCGATGAGTACATATACCAAAATGATATATTTTC 1974
QY  1700 CACTCCCGCATCGCATGTTGTTTGAAGACCCCTTGATATATGAGAAACTATATGCTTA 1759
    || || || || || || || || || || || || || || || || || || || || ||
DB  1975 CACTTCCAACTGAGATGCTTTTCAAAAACCTTTATGACCATACATGATGCTCTCA 2034
QY  1760 AGATTTGGAATCTCAACAAGTTCACGTGAGATGATGTCATTAACCTGCCAAGAGAG 1819
    || || || || || || || || || || || || || || || || || || || || ||
DB  2035 AATTTTGGAACTTCAACCAAGTATGAGAGGCGTATGTTGTCCTTCAAGTGTCAAGGGGCG 2094
QY  1820 GATGCTGTGTAAGACAGCCGCCAACCAATGCTTTTACATATCTCAAAACGATGACAT 1879
    || || || || || || || || || || || || || || || || || || || || ||
DB  2095 GGTGGATCTCAATTAATGCAAGAGTTTAGGGGCTTCCCTGAATGCTAACAAGCCAAATACCG 2154
QY  1880 CCAAAACTAACCCAAAGACATAGAAATGGCACAGTGGAGAAAAACCTATCTATTTGAAG 1939
    || || || || || || || || || || || || || || || || || || || || ||
DB  2155 GTACTGTCTATGTAAACCGAAGTTGAATGGATCAAAAGGAAAGAACATCTCATTTGGGTA 2214
QY  1940 GCGTTAAACCTTTGCGCTTACCTCTATCAAGCCAAAAACT---TATCTCTCCAAAGC 1996
    || || || || || || || || || || || || || || || || || || || || ||
DB  2215 AAGCAGAAAGATATATGATGCTACTTGAATCAAGCTGAGAGCTTTCTTGTATGATCTTA 2274
QY  1997 CCTTCAGATCTTGAATAGCTCTTGACCCATTCGAATTCGAGTCACTGATCTTTCAC 2056
    || || || || || || || || || || || || || || || || || || || || ||
DB  2275 AATCCGAACCAATTCAGTTACTATTAACACATCTAATTTGAGCTATACAGTTTGTTC 2334
QY  2057 CAGTGAACCAACTCATCCAACTTCTTACACTTTGGCCCAATTTGGGCTGGTGAACATGC 2116
    || || || || || || || || || || || || || || || || || || || || ||
DB  2335 CGGTTACAAA---GTTATGTGGCGGCATCAAAATTTGGTTCACCAAAATTTGGTGGACAAACATGT 2391
QY  2117 TTAACACTAGTGAGACCATCATCTGTGTGACTATGACGATGACCTAAGCTCAGTCGAGA 2176
    || || || || || || || || || || || || || || || || || || || || ||
DB  2392 TCAACAGTGCTGGAACAGTATATGATTTGGAAATATGTTG-----GAAATGCTGCAAGA 2445
QY  2177 TTGCTGTCAAGGGTGTGTGATGATGCGAGTATTTGATCGAAAAACCAAGGCTTGTG 2236
    || || || || || || || || || || || || || || || || || || || || ||
DB  2446 TTAAGGATTAAGGTGTGTGAGATTTCTTGTCTTATTTAGTAAATCACCAAGAGTTTC 2505
QY  2237 GTATTGATGGGAGGATGTGTTGGTCAAGTATGATCAGGACCAAAATGTGGTGTTCAG 2296
    || || || || || || || || || || || || || || || || || || || || ||
DB  2506 AGTTGATGATGTTGTGAGTGAATTTTGAAGTGGCTAGGTGATGGAATTTGTGTCAATG 2565
QY  2297 TGCCATGGCCAAATGATTTCTT 2317
    || || || || || || || || || || || || || || || || || || || || ||
DB  2566 TTCTTGAATGAAGAGCCTT 2586
    || || || || || || || || || || || || || || || || || || || || ||

```

Search completed: April 3, 2003, 18:19:27
Job time : 7057 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

Run on: April 3, 2003, 12:29:40 ; Search time 592 Seconds
(without alignments)
8947.128 Million cell updates/sec

Title:	US-09-675-208-4_COPY_56_2407
Perfect score:	2352
Sequence:	1 ATGGCTCCAGTTTAAAAA.....CGTTATCGACTACTTCTTT 2352

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

```
Searched:      2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters:  4370478
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

Database : N_Geneseq_101002:1

1:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1980.DAT.*
2:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1981.DAT.*
3:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1982.DAT.*
4:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1983.DAT.*
5:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1984.DAT.*
6:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1985.DAT.*
7:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1986.DAT.*
8:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1987.DAT.*
9:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1988.DAT.*
10:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1989.DAT.*
11:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1990.DAT.*
12:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1991.DAT.*
13:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1992.DAT.*
14:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1993.DAT.*
15:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1994.DAT.*
16:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1995.DAT.*
17:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1996.DAT.*
18:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1997.DAT.*
19:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1998.DAT.*
20:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA1999.DAT.*
21:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA2000.DAT.*
22:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA2001A.DAT.*
23:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA2001B.DAT.*
24:	/SID52/gcgcdatA/genseq/genseqn-emb1/NA2002.DAT.*

Pred. No. Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Match	Score	Length	DB	ID	Description
1	2352	100.0	2517	19	AAV22250	Cucumber raffinose
2	2352	100.0	2569	20	AAH61238	Cucumber raffinose
3	1041.4	44.3	2630	20	AA202029	Mustard raffinose
4	977.4	41.5	2498	19	AAV40801	Soybean raffinose
5	976.4	41.5	2498	22	AAH82523	Soybean raffinose
6	975.4	41.5	2690	20	AA282028	Sugarbeet raffinose
7	966.8	41.1	2343	22	AAH27438	Soybean polyuncle
8	966	41.1	2497	20	AAZ10002	Nucleic acid encodo
9	956.6	40.7	2746	19	AAV40800	Broad bean raffin

Salmon
Sardines

10	956.6	40.7	2746	20	AA021000
11	948.6	40.3	2524	21	AAAD0335
12	886.6	37.7	1762	19	AAVA0802
13	886.6	37.7	1762	20	AAAZ0103
14	866.4	36.8	1762	20	AAAZ2020
15	303.8	11.9	928	20	AAK61239
16	250.8	12.7	2780	20	AAAD0033
17	243.2	10.3	2868	21	AAAD0034
18	240.8	10.2	2842	21	AAAD0033
19	220.8	9.4	2653	21	AAAD0038
20	191.8	8.2	2668	21	AAAD0037
21	184	7.8	3060	21	AAAD0033
22	174.4	7.4	1816	21	AAAD0031
23	169.6	7.2	983	19	AAVA0803
24	159.6	7.2	993	20	AAAZ1004
25	114	4.8	822	21	AAAC4201
26	113.2	4.8	1036	24	ABNB8120
27	90.8	3.9	253	20	AAK61259
28	75.4	3.2	296	24	ABL73625
29	50.4	2.1	636	22	AAH87794
30	48.2	2.0	540	22	AAH87795
31	41.6	1.8	29255	23	AAAS9516
32	39.4	1.7	149671	24	ABK84177
33	39.2	1.7	1417	24	ABK11710
34	39.2	1.7	3408	22	AAI93894
35	39.2	1.7	3408	22	AAH18536
36	39.2	1.7	66566	21	AAAS3430
37	38.6	1.6	189	24	ABN79093
38	38.8	1.6	4159	22	AAH24055
39	38.2	1.6	28046	24	ABK65837
40	38	1.6	911	22	AAAD0334
41	38	1.6	911	22	AAAD0332
42	37.8	1.6	31422	21	AAAZ2020
43	37.8	1.6	31422	22	AAH79278
44	37.4	1.6	1664	22	AAAS3571
45	37.4	1.6	1664	22	AAAS3572

ALIGNMENTS

RESULT 1
AAV22250
ID AAV22250 standard: CDNA to mRNA; 2517 BP.

AC AAV22250;

DT 06-JUL-1998 (first entry)

Cucumber raffinose synthase cDNA.

KW Cucumber; raffinose synthase; sucrose; galactinol; ds

OS Cucumis sativus

Key	Location/qualifiers
FH	
EC	

```

/*tag= a
/next=

```

XX
DN 7510064073-A

XX 07-APR-1999

XX 28-APR-1997. 97TB-0111124
PF

XX 26-JUL-1996. 96JP-0198079
PB

PR 26-APR-1996; 96JF-010/682;
XX

XX PA (AOLN) AJJNOMOLU KK.

DR WP1: 1998-264858/24.
DB P-PSDB: AAW53570

XX Raffinose synthase gene - useful for preparation of raffinose in
PT transformed plant
XX
XX
PS Claim 7, Pages 17-20; 26pp; Japanese.
XX
CC The present sequence encodes cucumber raffinose synthase, which
CC forms raffinose from sucrose and galactinol, has an optimum pH of
CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
CC iodoacetamide, N-ethylmaleimide and myoinositol.
XX
SQ Sequence 2517 BP; 613 A; 524 C; 642 G; 738 T; 0 other;
Query Match 100.0%; Score 2352; DB 19; Length 2517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTCTAGTTTAAANAATGGTGGCTCCCAACGATGTTTCATTTGATGGCTTAAATGAC 60
DB ATGGCTCTAGTTTAAANAATGGTGGCTCCCAACGATGTTTCATTTGATGGCTTAAATGAC 115
QY 61 ATGTGTCACCGTTTGCATTCAGATTCGATTCACGTGACAGCTGATTCGTTTCG 120
DB ATGTGTCACCGTTTGCATTCAGATTCGATTCACGTGACAGCTGATTCGTTTCG 175
QY 121 TCCGATGTTCTGAGAACATTTGTTGCTTCCCTTCCGTACACTTCGATAGCAAGTCC 180
DB TCCGATGTTCTGAGAACATTTGTTGCTTCCCTTCCGTACACTTCGATAGCAAGTCC 235
QY 176 TCCGATGTTCTGAGAACATTTGTTGCTTCCCTTCCGTACACTTCGATAGCAAGTCC 235
QY 181 CCGGTTTCGGTTGGTTGCTTTGTTGGATTGACGCGTCGGAACCTGATAGCCGACATGTT 240
DB CCGGTTTCGGTTGGTTGCTTTGTTGGATTGACGCGTCGGAACCTGATAGCCGACATGTT 295
QY 241 GTTTCGATTGGGAAGCTGAAGGATATTCGTTATACAGATTTTTCAGGTTTAAAGTTTG 300
DB GTTTCGATTGGGAAGCTGAAGGATATTCGTTATACAGATTTTTCAGGTTTAAAGTTTG 355
QY 296 GTTTCGATTGGGAAGCTGAAGGATATTCGTTATACAGATTTTTCAGGTTTAAAGTTTG 355
QY 301 TGGACTACACACTGGGTTGGTGGAAATGGTGGGATCTTGAATCGAGAGCTCAGATTGTG 360
DB TGGACTACACACTGGGTTGGTGGAAATGGTGGGATCTTGAATCGAGAGCTCAGATTGTG 415
QY 356 TGGACTACACACTGGGTTGGTGGAAATGGTGGGATCTTGAATCGAGAGCTCAGATTGTG 415
QY 361 ATCCCTTGAGAAGTCAGATTCTGGTGCAGCGTATGTTTCTCTTCGATTCGTTGAGGA 420
DB ATCCCTTGAGAAGTCAGATTCTGGTGCAGCGTATGTTTCTCTTCGATTCGTTGAGGA 475
QY 416 ATCCCTTGAGAAGTCAGATTCTGGTGCAGCGTATGTTTCTCTTCGATTCGTTGAGGA 475
QY 421 CCGTTCCGAACCTCGATTCAGCTGGGATGATGACTTGTGATGTTTGTGCGAGAT 480
DB CCGTTCCGAACCTCGATTCAGCTGGGATGATGACTTGTGATGTTTGTGCGAGAT 535
QY 476 CCGTTCCGAACCTCGATTCAGCTGGGATGATGACTTGTGATGTTTGTGCGAGAT 535
QY 481 GGTTCGTCGAAGATGTTGATGATCGATTCGGAAGTATGTTGATTCATGCTGATGAT 540
DB GGTTCGTCGAAGATGTTGATGATCGATTCGGAAGTATGTTGATTCATGCTGATGAT 595
QY 536 GGTTCGTCGAAGATGTTGATGATCGATTCGGAAGTATGTTGATTCATGCTGATGAT 595
QY 541 GATTCGTTTGCACTTGTCTTAAAGAGCGATGAAGATCGTAGAGACCATCTTGGAACTTTT 600
DB GATTCGTTTGCACTTGTCTTAAAGAGCGATGAAGATCGTAGAGACCATCTTGGAACTTTT 655
QY 596 GATTCGTTTGCACTTGTCTTAAAGAGCGATGAAGATCGTAGAGACCATCTTGGAACTTTT 655
QY 601 GCCTTGTGAGAGAGAAGCTCCACAGGTATCGTGGACAATTTGGTGGTGGACAGTGG 660
DB GCCTTGTGAGAGAGAAGCTCCACAGGTATCGTGGACAATTTGGTGGTGGACAGTGG 715
QY 656 GCCTTGTGAGAGAGAAGCTCCACAGGTATCGTGGACAATTTGGTGGTGGACAGTGG 715
QY 661 GACGGTCTTTTACCTAAACGTTTCATCCACAGGGCGTAAATGAAGGCGTGAGGATCTCGTC 720
DB GACGGTCTTTTACCTAAACGTTTCATCCACAGGGCGTAAATGAAGGCGTGAGGATCTCGTC 775
QY 716 GACGGTCTTTTACCTAAACGTTTCATCCACAGGGCGTAAATGAAGGCGTGAGGATCTCGTC 775
QY 721 GACGGCGGTTTGTCTCCCGGTTTAACTCTTAATCGACGATGTTTGGCAATCCATCGGACAC 780
DB GACGGCGGTTTGTCTCCCGGTTTAACTCTTAATCGACGATGTTTGGCAATCCATCGGACAC 835
QY 776 GACGGCGGTTTGTCTCCCGGTTTAACTCTTAATCGACGATGTTTGGCAATCCATCGGACAC 835
QY 781 GATTTCGATCCCATCAACCAAGAGATGAACCAACCGTGGCGGCGACGAATGCCCC 840
DB GATTTCGATCCCATCAACCAAGAGATGAACCAACCGTGGCGGCGACGAATGCCCC 840

DB 836 GATTTCGATCCCATCAACCAAGAGATGAACCAACCGTGGCGGCGACGAATGCCCC 895
QY 841 TGGCGTCTTTTGAATTTCCACAGAGAAATTCAAAATTCCTGGATAGTCATCCCAAGGCC 900
DB TGGCGTCTTTTGAATTTCCACAGAGAAATTCAAAATTCCTGGATAGTCATCCCAAGGCC 955
QY 901 ACCGGCCCCGAGCCCGCCAGAAAGGGGATGAAGCCGTTTATAGATGACTCAAGAGAGAG 960
DB ACCGGCCCCGAGCCCGCCAGAAAGGGGATGAAGCCGTTTATAGATGACTCAAGAGAGAG 1015
QY 956 ACCGGCCCCGAGCCCGCCAGAAAGGGGATGAAGCCGTTTATAGATGACTCAAGAGAGAG 1015
QY 961 TTTTAACACTGTGACATGTTTATGTTGGCATGCTTTGTGTGATATTTGGGTGGCCTT 1020
DB TTTTAACACTGTGACATGTTTATGTTGGCATGCTTTGTGTGATATTTGGGTGGCCTT 1075
QY 1016 TTTTAACACTGTGACATGTTTATGTTGGCATGCTTTGTGTGATATTTGGGTGGCCTT 1075
QY 1021 CGCCCGCAGGTGCTGCTGCTTGGCTTCGAGGACGCTGTATTCAGCAGTGGCTTCCACAGGG 1080
DB CGCCCGCAGGTGCTGCTGCTTGGCTTCCGAGGACGCTGTATTCAGCAGTGGCTTCCACAGGG 1135
QY 1076 CGCCCGCAGGTGCTGCTGCTTGGCTTCCGAGGACGCTGTATTCAGCAGTGGCTTCCACAGGG 1135
QY 1081 CTGACAGATGACGATGAGAGATTTTGGCGGTGATTAAGATTGTTCTTCATTAAGTGGGCTG 1140
DB CTGACAGATGACGATGAGAGATTTTGGCGGTGATTAAGATTGTTCTTCATTAAGTGGGCTG 1195
QY 1136 CTGACAGATGACGATGAGAGATTTTGGCGGTGATTAAGATTGTTCTTCATTAAGTGGGCTG 1195
QY 1141 CTGCGCGCGGAGAGAGCTGAGGATGTACGAAGGACTTCATGCTCATTTTGGAAAAAGTT 1200
DB CTGCGCGCGGAGAGAGCTGAGGATGTACGAAGGACTTCATGCTCATTTTGGAAAAAGTT 1255
QY 1196 GTCCCGCCGAGAGAGCTGAGGATGTACGAAGGACTTCATGCTCATTTTGGAAAAAGTT 1255
QY 1201 GGGATGACGAGGTTTAAAGATTGACGTTATCCACTTATGGAGATGTTGTGTAAGACTAT 1260
DB GGGATGACGAGGTTTAAAGATTGACGTTATCCACTTATGGAGATGTTGTGTAAGACTAT 1315
QY 1256 GGGATGACGAGGTTTAAAGATTGACGTTATCCACTTATGGAGATGTTGTGTAAGACTAT 1315
QY 1261 GCGAGGAGAGTGGATTGTGCAAAAGCATATTAACAAGCAATGACCAATCAATTAATAA 1320
DB GCGAGGAGAGTGGATTGTGCAAAAGCATATTAACAAGCAATGACCAATCAATTAATAA 1375
QY 1316 GCGAGGAGAGTGGATTGTGCAAAAGCATATTAACAAGCAATGACCAATCAATTAATAA 1375
QY 1321 CATTTTAAAGAAATGAGCACTATTCGAAATGGAACATTTGAAACACTTCATGTTCCCTT 1380
DB CATTTTAAAGAAATGAGCACTATTCGAAATGGAACATTTGAAACACTTCATGTTCCCTT 1435
QY 1376 CATTTTAAAGAAATGAGCACTATTCGAAATGGAACATTTGAAACACTTCATGTTCCCTT 1435
QY 1381 GGCACGAGAGCTATCTCTCTGCTGCTGTTGGTGAATGATGATTTGGTGCAGGACCCCTT 1440
DB GGCACGAGAGCTATCTCTCTGCTGCTGTTGGTGAATGATGATTTGGTGCAGGACCCCTT 1495
QY 1436 GGCACGAGAGCTATCTCTCTGCTGCTGTTGGTGAATGATGATTTGGTGCAGGACCCCTT 1495
QY 1441 GGTGATCCAAAGGTTACGTTTGGCTCCAAAGATGTCACATGTTTATTTGTCCAACGAC 1500
DB GGTGATCCAAAGGTTACGTTTGGCTCCAAAGATGTCACATGTTTATTTGTCCAACGAC 1555
QY 1496 GGTGATCCAAAGGTTACGTTTGGCTCCAAAGATGTCACATGTTTATTTGTCCAACGAC 1555
QY 1501 AGCTTGTGATGGGAGACTTCATCCACCCCTGAGTGGAGATGTTGCCAATCCACCCACT 1560
DB AGCTTGTGATGGGAGACTTCATCCACCCCTGAGTGGAGATGTTGCCAATCCACCCACT 1615
QY 1556 AGCTTGTGATGGGAGACTTCATCCACCCCTGAGTGGAGATGTTGCCAATCCACCCACT 1615
QY 1561 TGTGCCGCTTCATGCTGCTCTCGAGGCACTCTGTGTGGCCGCACTATGTTAGTGAT 1620
DB TGTGCCGCTTCATGCTGCTCTCGAGGCACTCTGTGTGGCCGCACTATGTTAGTGAT 1675
QY 1616 TGTGCCGCTTCATGCTGCTCTCGAGGCACTCTGTGTGGCCGCACTATGTTAGTGAT 1675
QY 1621 TGTGTGGGAAACATTAACCTTTGATCTTGAAGAAAACTAGTCTTCCTATGATGATGATC 1680
DB TGTGTGGGAAACATTAACCTTTGATCTTGAAGAAAACTAGTCTTCCTATGATGATGATC 1735
QY 1676 TGTGTGGGAAACATTAACCTTTGATCTTGAAGAAAACTAGTCTTCCTATGATGATGATC 1735
QY 1681 CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
DB CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1795
QY 1736 CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1795
QY 1741 AATGGAACACTATGCTTATGATTTGGAATCTCAACAATGTTACTGAGAGTGAATGTTGTGA 1800
DB AATGGAACACTATGCTTATGATTTGGAATCTCAACAATGTTACTGAGAGTGAATGTTGTGA 1855
QY 1796 AATGGAACACTATGCTTATGATTTGGAATCTCAACAATGTTACTGAGAGTGAATGTTGTGA 1855
QY 1801 TTTCACTGCGCAAGAGAGAGATGTTGCTGTGAGACAGCCGCAACCAATGCTTTTCACAA 1860
DB TTTCACTGCGCAAGAGAGAGATGTTGCTGTGAGACAGCCGCAACCAATGCTTTTCACAA 1915
QY 1856 TTTCACTGCGCAAGAGAGAGATGTTGCTGTGAGACAGCCGCAACCAATGCTTTTCACAA 1915
QY 1861 TACTCAAAAACGATGATCCCAAAAACCTAATCCCAAAAAGCATATGATGATGATGATGATGAT 1920
DB TACTCAAAAACGATGATCCCAAAAACCTAATCCCAAAAAGCATATGATGATGATGATGATGAT 1975

OY 1921 AACCTATCTCTATTGAAAGCGCTTAAACCTTGGCTTACCTTATCAAGCCAAAAA 1980
|||||
DB 1976 AACCTATCTCTATTGAAAGCGCTTAAACCTTGGCTTACCTTATCAAGCCAAAAA 2035
OY 1981 CTTATCTCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCCATTCGATTCGAG 2040
CTTATCTCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCCATTCGATTCGAG
DB 2036 CTTATCTCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCCATTCGATTCGAG 2095
OY 2041 CTCATCAGCTGTTTCAAGCTGAGCCAAACCTCAACCTTCTCTACCTTGGCCCAATT 2100
|||||
DB 2096 CTCATCAGCTGTTTCAAGCTGAGCCAAACCTCAACCTTCTCTACCTTGGCCCAATT 2155
OY 2101 GGGCTGTGTAACATCTTAACTAGTGGAGCCATCAATCTGTGAGATGACATGAC 2160
2156 GGGCTGTGTAACATCTTAACTAGTGGAGCCATCAATCTGTGAGATGACATGAC 2215
OY 2161 CTAAGCTCAGTCGAGATGTTGTCTCAAGGGTGTGTGAGATGCGAGATTTGCAATCGAAA 2220
|||||
DB 2216 CTAAGCTCAGTCGAGATGTTGTCTCAAGGGTGTGTGAGATGCGAGATTTGCAATCGAAA 2275
OY 2221 AAACCAAGGCGTGTCTGATTTGATGGGAGAGATGTTGGTTCAATGATGATGAGACCAA 2280
|||||
DB 2276 AAACCAAGGCGTGTCTGATTTGATGGGAGAGATGTTGGTTCAATGATGATGAGACCAA 2335
OY 2281 ATGCTGTGTGTTCAAGTGCATGCGCAATTGATCTTCTCGGTGCGATTTGCGTTATC 2340
|||||
DB 2336 ATGCTGTGTGTTCAAGTGCATGCGCAATTGATCTTCTCGGTGCGATTTGCGTTATC 2395
OY 2341 GAGTACTTGTGTT 2352
|||||
DB 2396 GAGTACTTGTGTT 2407
RESULT 2
AAK61238
ID AAK61238 standard: cDNA to mRNA; 2569 BP.
XX
AC AAK61238:
XX
DT 29-JUL-1999 (first entry)
XX
DE Cucumber raffinose synthase encoding cDNA.
XX
KM Raffinose synthase; sucrose; galactinol; ds.
XX
OS Cucumis sativus.
XX
FH Key Location/Qualifiers
FT CDS 56..2410
FT /tag= a
XX
PM JPI1123080-A.
XX
PD 11-MAY-1999.
XX
PE 24-OCT-1997; 97JP-0292969.
XX
PR 24-OCT-1997; 97JP-0292969.
XX
PA (AJIN) AJINOMOTO KK.
XX
DR MPI: 1999-340516/29.
XX
DR P-PSDB: AAY17417.
XX
PT New raffinose synthase gene - for production of raffinose from
PT sucrose and galactinol
XX
PS Claim 10; Page 22-25; 37pp; Japanese.
XX
CC The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of

CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence encodes raffinose
CC synthase from cucumber.
XX
SO Sequence 2569 BP; 641 A; 527 C; 643 G; 755 T; 3 other:
Query Match 100.0%; Score 2352; DB 20; Length 2569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGGCTCTAGTATTAAAAATGGTGGCTTCACAGCTAGTTTCAATTTGATGCTTAAATGAC 60
DB 56 ATGGCTCTAGTATTAAAAATGGTGGCTTCACAGCTAGTTTCAATTTGATGCTTAAATGAC 115
OY 61 ATGTCGACCGTTTGGCATGAGGAGATGCGATTTTACGTGTAACGGTCATTCGTTGCG 120
DB 116 ATGTCGACCGTTTGGCATGAGGAGATGCGATTTTACGTGTAACGGTCATTCGTTGCG 175
OY 121 TCCGATGTTTCTGAGAACATTTGCTTCTCTCTCCGTACCTTGCATTAAGACAAATGCC 180
DB 176 TCCGATGTTTCTGAGAACATTTGCTTCTCTCTCCGTACCTTGCATTAAGACAAATGCC 235
OY 181 CCGGTTTGGGTTGTTGCTTTGTTGGATTGCAACGCGTGGAACTTGATAGCCGACATGTT 240
DB 236 CCGGTTTGGGTTGTTGCTTTGTTGGATTGCAACGCGTGGAACTTGATAGCCGACATGTT 295
OY 241 GTTTCGATTTGGGAACTGAGATATTCGGTTTATGAGATTTTGAAGTTTGAAGTTTGG 300
DB 296 GTTTCGATTTGGGAACTGAGATATTCGGTTTATGAGATTTTGAAGTTTGAAGTTTGG 355
OY 301 TGGACTACACACTGGGTTTGTGTCGAATGTTGGGATCTTGTGAATCGAGACTCAGATTGTG 360
DB 356 TGGACTACACACTGGGTTTGTGTCGAATGTTGGGATCTTGTGAATCGAGACTCAGATTGTG 415
OY 361 ATCTTTGAGAACTCAATTTCTGTCGACCGCTATGTTTTCCTTCCATGCTTGAAGGA 420
DB 416 ATCTTTGAGAACTCAATTTCTGTCGACCGCTATGTTTTCCTTCCATGCTTGAAGGA 475
OY 421 CCGTTCCGAACTGATTCAGCTGCGGAGTGAATGATTTGTCGATGTTTGTGTCAGAGT 480
DB 476 CCGTTCCGAACTGATTCAGCTGCGGAGTGAATGATTTGTCGATGTTTGTGTCAGAGT 535
OY 481 GGTTCGTCGAAAGTTTGTGATGATCATGTTCCGAACTGATTTGATCTTCAATCTGCTGAT 540
DB 536 GGTTCGTCGAAAGTTTGTGATGATCATGTTCCGAACTGATTTGATCTTCAATCTGCTGAT 595
OY 541 GATCCGTTTGGCACTTTGTTAAGAGGCGATGAAGATCTGTGAGACCCATCTTGAACCTTTT 600
DB 596 GATCCGTTTGGCACTTTGTTAAGAGGCGATGAAGATCTGTGAGACCCATCTTGAACCTTTT 655
OY 601 CGCTTGTTCGAGAGAAAGCTCCACAGGTATGTCGTCGCAAAATTTGGTTCGTCACAGTCG 660
DB 656 CGCTTGTTCGAGAGAAAGCTCCACAGGTATGTCGTCGCAAAATTTGGTTCGTCACAGTCG 715
OY 661 GACGCGTTTACCTTAACGCTTCATCCACAGGCGCTTAATGAAGGCGTGAAGCATCTGTC 720
DB 716 GACGCGTTTACCTTAACGCTTCATCCACAGGCGCTTAATGAAGGCGTGAAGCATCTGTC 775
OY 721 GACGCGGTTTGTCTCCGCGTTTACTCTTAATGACAGATGTTGGCAATCCATCCGACAC 780
DB 776 GACGCGGTTTGTCTCCGCGTTTACTCTTAATGACAGATGTTGGCAATCCATCCGACAC 835
OY 781 GATTCGGATCCCATCAGCAAGAAAGAAATGAACCAACCGTCGCGGAGCAAAATGCC 840
DB 836 GATTCGGATCCCATCAGCAAGAAAGAAATGAACCAACCGTCGCGGAGCAAAATGCC 895
OY 841 TGCCTCTTTTGAATTCAGAGAAATTAACAATTCCTGATAGCTGAATCCCAAGGCC 900
DB 896 TGCCTCTTTTGAATTCAGAGAAATTAACAATTCCTGATAGCTGAATCCCAAGGCC 955
OY 901 ACCGCGCCCGGAGCGCGCAGAGGAGATGAAGCGTTTATGATGAACCTCAAGAGAG 960
DB 956 ACCGCGCCCGGAGCGCGCAGAGGAGATGAAGCGTTTATGATGAACCTCAAGAGAG 1015

Oy	961	TTTAAAGCTGAGGACATGTTTATGTTTGGCAAGCTTTGTGGADATTTGGGATGGCCTT	102
Db	1016	TTTAAAGCTGAGGACATGTTTATGTTTGGCATGCTTTGTGGATATTGGGATGGCCTT	1075
Oy	1021	CGCCCGAGGTGCTGCTGGCTTGCCTGAGGCAAGCTGTGTATTCAGCCAGTCTTTCACCAAGG	1080
Db	1076	CGCCCGAGGAGTCCGTGCTTGCCTGAGGCAAGCTGTATTCAGCCAGTCTTTCACCAAGG	1135
Oy	1081	CTGAGATGACGATGAGGAGATTGGGGGGTGGATTAATGTTCTTCATTAAGTGGCGCTG	1140
Db	1136	CTGAGATGACGATGAGGAGATTGGGGGGTGGATTAATGTTCTTCATTAAGTGGCGCTG	1195
Oy	1141	GTCGCCGCGGAGAAAGGCTGAGGAGATGTACGAAGACCTTCATGCTCATTTGGAAAAAGTT	1200
Db	1196	GTCGCCGCGGAGAAAGGCTGAGGAGATGTACGAAGACCTTCATGCTCATTTGGAAAAAGTT	1255
Oy	1201	GGGATCCAGCGTGTTAAGATTGACGTTATCCACTATTGGAAATGTTGTGTAGACATAT	1260
Db	1256	GGGATCCAGCGTGTTAAGATTGACGTTATCCACTATTGGAAATGTTGTGTAGACATAT	1315
Oy	1261	GGAGGAGAGAGTGATTTGGCAAGAAGCATATACAAAGCATAGCCAAATCAATAATATAA	1320
Db	1316	GGAGGAGAGAGTGATTTGGCAAGAAGCATATACAAAGCATATACCAATCAATAATATAA	1375
Oy	1321	CATTTTAAAGAAATGGAGTCAATTGCAAGTATGCAACATTTGTAAAGCTCATGTTCCTT	1380
Db	1376	CATTTTAAAGAAATGGAGTCAATTGCAAGTATGCAACATTTGTAAAGCTCATGTTCCTT	1435
Oy	1381	GGCAGGAGAGTATCTCTGTGGTGGTGGTGGATGATCTTTGGTGGCAGGAGCCCTCT	1440
Db	1436	GGCAGGAGAGTATCTCTGTGGTGGTGGTGGATGATCTTTGGTGGCAGGAGCCCTCT	1495
Oy	1441	GGTGATCCAAACGGTACGTTTGGCTCCAAAGATGTACATGGTTCATTTGTGCCAAGAC	1500
Db	1496	GGTGATCCAAACGGTACGTTTGGCTCCAAAGATGTACATGGTTCATTTGTGCCAAGAC	1555
Oy	1501	AGCTTGTGAGTGGGAACTTCAATCCACCTGACTGGGATATGTTCATCCACCACCTT	1560
Db	1556	AGCTTGTGAGTGGGAACTTCAATCCACCTGACTGGGATATGTTCATCCACCACCTT	1615
Oy	1561	TGTGCCCGCTTCATGCTGCTGCCCTCCAGGCACTCTGATGGGCCCATGTATTTAGTAT	1620
Db	1616	TGTGCCCGCTTCATGCTGCTGCCCTCCAGGCACTCTGATGGGCCCATGTATTTAGTAT	1675
Oy	1621	TCTGTGGGAAAGCATTAATCTTTGATCTTTCTGAAAAAACTAGTCTTCTGTATGATCATC	1680
Db	1676	TCTGTGGGAAAGCATTAATCTTTGATCTTTCTGAAAAAACTAGTCTTCTGTATGATCATC	1735
Oy	1681	CTTGAGAGTGTAGTACATGCACTCCGACCTCGCGATGTTGTTTGTGAAGACCTTTGCAT	1740
Db	1736	CTTGAGAGTGTAGTACATGCACTCCGACCTCGCGATGTTGTTTGTGAAGACCTTTGCAT	1795
Oy	1741	AATGGAGAAACTATGCTTAAGATTTGGAATCTCAACAAATTCACTGAGTGTATGTGTCA	1800
Db	1796	AATGGAGAAACTATGCTTAAGATTTGGAATCTCAACAAATTCACTGAGTGTATGTGTCA	1855
Oy	1801	TTTCAACTGCCAAGGAGAGATGTGTCTGTAGACACGCGCCGAACCATGCTTTTCACAA	1860
Db	1856	TTTCAACTGCCAAGGAGAGATGTGTCTGTAGACACGCGCCGAACCATGCTTTTCACAA	1915
Oy	1861	TATCACAAAAGAGAGACATCAAAATTAACCCAAAAAGACATAGATGTGCACAGTGGAGAA	1920
Db	1916	TATCACAAAAGAGAGACATCAAAATTAACCCAAAAAGACATAGATGTGCACAGTGGAGAA	1975
Oy	1921	AACCCATCTCTATTTGAAGCGTTAAACCTTTGGCGTTTACCTCTATCAAGCCAAAAA	1980
Db	1976	AACCCATCTCTATTTGAAGCGTTAAACCTTTGGCGTTTACCTCTATCAAGCCAAAAA	2035
Oy	2036	CTTATCTCTCTCAAGGCTCTCTCAAGATCTTTGACATAGCTCTTGGACCACTTGGAAATTCGAG	2040
Db	2091	CTTATCTCTCTCAAGGCTCTCTCAAGATCTTTGACATAGCTCTTGGACCACTTGGAAATTCGAG	2095

Oy	2041	CTATCAGCTGTTTACCACAGACCAACATCATCAAACTTCTCTACATTTGCCCAAT	2100
Db	2096	CTCATCACTGTTTACCACAGACCAAACTCATCAAACTTCTCTACATTTGCCCAAT	2155
Oy	2101	GGGCTGGTGAACATGCTTAACTAGTGGAGCCATCCAACTCTGTGACTATGACGATAC	2160
Db	2156	GGGCTGGTGAACATGCTTAACTAGTGGAGCCATCCAACTCTGTGACTATGACGATAC	2215
Oy	2161	CTAAGCTCAGTCGAGATTGGTGTCAAAAGGCTGTGGAGAAATCGAGATATTTCATCGAAA	2220
Db	2216	CTAAGCTCAGTCGAGATTGGTGTCAAAAGGCTGTGGAGATCGAGATATTTCATCGAAA	2275
Oy	2221	AAACCAAGGCTTGCTATTTGAATGGGAGAGATGGTGGTTAAATATGATCAGACCAA	2280
Db	2276	AAACCAAGGCTTGCTATTTGAATGGGAGAGATGGTGGTTAAATATGATCAGACCAA	2335
Oy	2281	ATGTGTGTTGTTTCAAAGTCCATGGCCAAATTGATTTCTTCATCGGTGCATTTCCGTTATC	2340
Db	2336	ATGTGTGTTGTTTCAAAGTCCATGGCCAAATTGATTTCTTCATCGGTGCATTTCCGTTATC	2395
Oy	2341	GAGTACTTGTTT 2352	
Db	2396	GAGTACTTGTTT 2407	

RESULT 3	
AAZ20209	
ID	AAZ20209 standard; cDNA; 2690 BP.
XX	
AC	AAZ20209;
XX	
DT	17-JAN-2000 (first entry)
XX	
DE	Mustard raffinose synthase cDNA.
XX	
KW	Raffinose synthase; mustard; transgenic plant; ss.
XX	
OS	Brassica juncea.
XX	
FH	Key
FT	Location/Qualifiers
FT	134..2467
FT	/*tag=a
FT	/note="this region is specifically claimed in
FT	Claim 7"
XX	
PN	EP93643-A2.
XX	
PD	03-NOV-1999.
XX	
PF	27-APR-1999; 99EP-0107430.
XX	
PR	30-APR-1998; 98JP-0120550.
PR	30-APR-1998; 98JP-0120551.
PR	04-DEC-1998; 98JP-0345590.
PR	10-DEC-1998; 98JP-0351246.
XX	
PA	(SUMO) SUMITOMO CHEM CO LTD.
XX	
PI	Watanabe E, Oeda K;
XX	
DR	WPI; 1999-593144/51.
XX	
PT	P-PSDB; AAY32074.
XX	
PT	New sense and antisense genes, useful for altering the level of
XX	raffinose in food plants -
PS	Claim 7; Page 32-36; 55pp; English.
XX	
CC	This is the nucleotide sequence of a mustard cDNA clone coding
CC	for raffinose synthase (see AAY32074), a protein which can bind a
CC	D-galactosyl group through an alpha(1-6) bond to the hydroxy group
CC	attached to the carbon atom at the 6-position of the D-glucose
CC	residue in a sucrose molecule to form raffinose. The cDNA was

CC Isolated from mustard (Brassica juncea) leaf cDNA by PCR. Probes
CC or primers generated from plant raffinose synthase genes (see
CC AA20207-10) may be used to obtain other raffinose synthase genes
CC by labeled detection or amplification (claimed). These genes may
CC be used to control the levels of raffinose produced in plants.
CC Antisense genes can be used to knock out existing gene activity.
CC and sense genes to increase the level of gene activity. The
CC resulting transgenic plants may be used as a food source to alter
CC the growing conditions for gut enterobacteria, providing general
CC health advantages. The sense and antisense genes may also be used
CC in gene and phenotypic analysis of plants and for the selection of
CC plant clones with the desired characteristics with respect to
CC raffinose content.

SO Sequence 2690 BP: 648 A; 608 C; 668 G; 765 T; 1 other:

Query Match 44.3%; Score 1041.4; DB 20; Length 2690;
Best Local Similarity 68.3%; Pred. No. 0;
Matches 1530; Conservative 0; Mismatches 676; Indels 33; Gaps 5;

QY 80 TCAGAGGATCGGATTTCACTGTGAACGGGTCATTGCTTCTGCGGATGTCGGAACA 139
DB 210 TAGAGGATTCGATCTCCATGCAATGATGACGTTGCTTAACCGATGATGACGTTAACG 269
QY 140 TTGTGCTTCTCCTCTCCGTAC-----ACTTCGATAGACAAGTCCCGGTTTCG 189
DB 270 TCAGCTGTCACTGCTTCACTTACCTAGTACGACAAGACGAGACCGGTTGACGGCTCCG 329
QY 190 GTTGGTGTGTTGTGGATTCACCGCTGC--GAACCTATAGCCGACATGTTGTTTGA 247
DB 330 CTGGTTCATTTCATCGGGTTTAATCTCGACGGTGAGCCAGGAAGCCGCGCTGCTCA 389
QY 248 TTGGGAAGCTGAAGATTTATCGTTTATGATTTTAAAGTTTAAAGTTTGGACTA 307
DB 390 TCGGTAAGTCAAGGATATTCATTCATGACATATTCGTTTCAAGTTTGGTGACTA 449
QY 308 CACACTGGGTGTCGAAATGCTGGGATCTTGAATCGAGACTCAGATTTGATCTTGG 367
DB 450 CTCACGTGGGTTCGTTCCAAAGGATCCGACATCGAGAACGACCCGATCATCATCTCG 509
QY 368 AAGATCAGATTCGTGTGACCGCTATGTTTCTTCCGATCGTTGAGGACGCTTC 427
DB 510 AAGACTCCGGTCTGCTCTTATGTTCTTCTGCGCTTCTTGAAGGCTCTTTC 569
QY 428 GAACCTGATAGGCTGGGATGATGACTTTGTCGATGTTTGTGTCGAGAGTGTCT 487
DB 570 GTTTCATCTTTAAGCTGGGGAAGACGATGACGTGGGTTGTGTGCAATCCGGTCA 629
QY 488 CGAAGTGTGATGATGATCTCGAAGTATGTTGATCTTCATGCTGTGATGATCCGT 547
DB 630 CCCAGGTGACCGGGTCCGAGTTTCTGCAAGTTGTGTATGTTCAACGCGGAGATCCGT 689
QY 548 TTGCACTTTTAAAGAGGATGAAAGATGCTGAGGACCATCTTGGAACTTTTCCGTTGT 607
DB 690 TCAAGTCTGTGAAGACGCGATGAAAGTGTGAGGTTTATATGAACACCTTCAAGCTCT 749
QY 608 TGGAGAGAAAGCTCCACGAGTATGCTGAGACAAATTCGGTTGGTGCAGCTGGAGACGT 667
DB 750 TGGAGAGAAAGRCGCGCGGGAATGCTGATAGTTTCGGGTGTGTCACGTGGGAATCGCT 809
QY 668 TTATACCTAAGCTTCAATCACAAGGCGCTAATAGAAAGCTGAGGCAATCTGTCGACGCG 727
DB 810 TTATATTGACGGTGAACCTTACGAGGATCATTAAGGTTTAAAGTGTCTGTCGACGCGTG 869
QY 728 GTTGTCCCTCCGTTTATGCTAATGAGAGATGTTGGAAATTCATCCGACACAGATCCG 787
DB 870 GTTGTCCGCGGGATGTTGCTAATGACGAGCGTTGCAATCGATTTGACATGATCCG 929
QY 788 ATCCCATCAACAAGAAATGAACAACAACCGTCCGCGGCAAAATGCCCTCGCTC 847
DB 930 ATGATATGATGTGAAGGATGAGTTTACCGTCCGCGGGAACAATGCCCTTGAAGC 989
QY 848 TTTTGAATTTCCAAAGAAATTAACAATTTCCGTACTACTCAATCCCAAGGCCACCGGCC 907

DB 990 TTCTGAATTTCAAGAACTTCAAGTTTCAGAGACTACGTCCTCCGAAGACA----- 1043
QY 908 CCCGAGCCGCCAGAAAGGAGATGAAGCGCTTATATGATGAATCAAAAGAGATTAA 967
DB 1044 -----AAACCAAGTCCGGATGAAGCTTTTCTGATAGATTTGAAGAAATTTCTCA 1097
QY 968 CTGTGAGCATTTTATGTTTGGATGCTTGTGATATTTGGGATGGCTTCGCCCGC 1027
DB 1098 CCGTGTATTACATCTACGCTGTGGACGCGCTTTGGGGCTACGTGGGTGTCTTCCGCC 1157
QY 1028 AGTGCCTGGCTTGGCTGAGCAGCTGTGATTCAGCCAGTGTCTTACACAGGCTGCAGA 1087
DB 1158 GAGCTCTACTCTTCGCCCTCAACTATTTGTCGGCCAGAGCTTCGCCGGGCTTAAAT 1217
QY 1088 TCAGCATGAGAGATTTGGCGGTGGATTAAGATGTTCTTATTAAGTCCGGGTGTCGCC 1147
DB 1218 TCAGCATGCAAGATCTCCGCTGTGATGAATGTGCAATCCGGAATTCGATTCGCTCGC 1277
QY 1148 CGAAGAGCTGAGGATGATGTACGAAGCACTTCATGCTCATTTGGAAAAAGTTGGGATCG 1207
DB 1278 CGGACATGCGGAATGATGTTTACGAAGGCTTCTACTCTCATCTTCAAAACGTCGGTATTG 1337
QY 1208 ACGGTATTAAGATGACGTTATCCACCTAATTTGAGATGTTGTGAAGACATGAGGGA 1267
DB 1338 ACGGCTTAAAGTTGACGTCATCCACATATGAGATGTTGTGAGAAATATGCGGGA 1397
QY 1268 GAGTGGATTTGGCAAAAGCATATTCAAAAGCAATGACCAATCAATTAATTAACATTTTA 1327
DB 1398 GAGTGAAGCTTGAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAG 1457
QY 1328 AAGGAATGAGATCATTTGCAAGATGATGAATGATTAAGCACTTCAATGTTCTTGGACAG 1387
DB 1458 ACGTGAACGCGCTTATCGGTACATGAGACACTGATATGATTTTCAATGTTCTTGGACAG 1517
QY 1388 AAGCTATCTCTGTGCTGCTGTTGATGATGACTTTTGTGTCAGGACCCCTCTGTGATC 1447
DB 1518 AAGCATCTCTTAAGTGTGCTGCTGATGATGACTTTTGTGTCAGGATCATCAGCGACA 1577
QY 1448 CAAGCGTACGTTTGGCTCCCAAGGATGTCACATGTTGATGTTGTCGCAAGGACGACTGT 1507
DB 1578 TAAAGGACATATTTGGCTGCAAGAGATGCAATGCTACAGTCTTCAACACGCTT 1637
QY 1508 GGATGGGCACTTCAATCCACCCCTGACTGGATGATGTTCCAAATCCACCCCTTGGCCG 1567
DB 1638 GGATGGGAATTTCAATCCAGCCTGATTTGGACATGTTTCAAGTCTTGTGTG 1697
QY 1568 CCTTCATGCTCTCTGAGGCAATCTGTGATGCTGATGCTATGTTTACGATCTGTGG 1627
DB 1698 AGTTCATGCTCTCTGTCGATCTCGGATCTCGGTCGATTTACATCAGCAGATGTGTGG 1757
QY 1628 GAAACGATTAAGTATGCTTCTGAAAAGAACTGATGCTTCTGATGATGATGATGATG 1687
DB 1758 GCGAGCAGATTTGATGATCTTGAAGCGCTGCTGTGCTACGATGATGATGATGATG 1817
QY 1688 GTGAGTACTATGCTCCGACCTCGGATGTTGTTTGAAGACCTTTCATTAATGAG 1747
DB 1818 GTGAGCAGATGCTACCTCCCACTCGTGAAGCTGCTTTTGAACACCTCTTCAATGATGGA 1877
QY 1748 AAGCTATGCTTAAAGATTGGAATCTCAACAGTTCATGAGATGATGATGATGATGATG 1807
DB 1878 AAGCATGCTCAAGATTTGGAAGCTTGAACAGTACCTGGAATTTATGAGCATTTCAACT 1937
QY 1808 GCGAAGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1867
DB 1938 GCGAAGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1997
QY 1868 AAGGATGATCAATCAAAATCAACCCCAAAAGACATAGATGACACAGTGGAGAAACCTTA 1927
DB 1998 ACAGCTTAACCGCCCAACAATAATCTTAAGGAGCTTGAATGAACAGTGGAGAAACCCCA 2057
QY 1928 TCTCTATTGAAGCGTTAAACCTTTGGCTTTTACCTCTATCAAGGCAAAAACCTTATCC 1987


```

Db 958 ACAGATTGTAGTGGAAAGATTCTG-----AGAAAGGATGGTGCCTT 1002
Oy 939 TTTAATATGAACCAAGAGAGATTTAAGACTGTGAGACATGTTTNGTTTGGACATGCTTT 998
Db 1003 TTTAAGGCACTTGAAGGAACATTTTAGAGCGTGGAGCGTGTGTGTGTGTGTGTGTGTGT 1062
Oy 999 GTGTGGATTTATGGGGTGGCCCTTCGCCGAGGTGCTGCTTGCCTGAGGACAGCTGTGAT 1058
Db 1063 TTTGTGGATTTTGGGGTGGCGGTGTAGACCCAGAGTTCGGGCAATGCCCAAGCTTAAGTGT 1122
Oy 1059 TCAGCCAGTCTTTTCAACGAGGCTGCAGATGACGATGAGGATTTTGGCGGTGATTAAGAT 1118
Db 1123 CACTCCGAAGCTGTCCAAATGAGCTAAATTTGACAAATGAAGGATTTAAGCGGTGATTAAGAT 1182
Oy 1119 TTTTCTTCAATAGGTTCGGGCTGCTCCCGGAGAGAGGTGAGAGATGTAGAGAGACT 1178
Db 1183 CCTCAGTAAAGGAGTTGGACTGTGTGCACACACCTGGCTCACCTTTTGTATGACAGGGGCT 1242
Oy 1179 TCATGCTCATTTTGAAGAAAGTTGGATCGACGCTGTAAAGATTGACGTTATCCACTATT 1238
Db 1243 CCAGTCCGTTTGAATGTGGGGTATTGAGCGTGTAAAGTTGACGTTATACACTGCT 1302
Oy 1239 GGAGATGTTGTGAAGACTATGAGAGGAGAGTGGATTTGGCAAAAGCATATTACAAAC 1298
Db 1303 CGAGATGCTATCCGAGGATACGGTGGCGTGTGAGCTAGCCAAAGCTTATTACAAAC 1362
Oy 1299 AATGACCAAAATCAATAAATTAATTTAAGAAATGAGATTTGCAAGTATGAGAAC 1358
Db 1363 GCTCAGTCTGTTGGTGAAGACATTTTCAAGCAATGGGCTATGTCATGTCAGATGAGAGA 1422
Oy 1359 TTGTAAAGCATTTATGCTTCCTTGGCAGCGAAGCTATCTCTCTTGTGCTGTGTGTGATGA 1418
Db 1423 TTGTATATGACTTTTCTCTCTGTATACGAAACCAATAGCCCTTGGGCGGTGAGAGATGA 1482
Oy 1419 CTTTGTGTGACGAGACCCCTGTGTGATCAAAAGCTTGTGCTTCCAAAGATGTCA 1478
Db 1483 TTTTGTGTGACTGTATCCCTGTGAGATCCAAATGTCACATTTGCTTCCAAAGGCTCA 1542
Oy 1479 CATGGTTCATTGTGGCCAGACAGCTGTGTGATGGGGAAGCTTCCACCTCGACTGGGA 1538
Db 1543 CATGTGCACTGTGCTTACAAACGCTGTGTGATGGGGAATTTTATTCACCGGATTTGGGA 1602
Oy 1539 TATGTTCCATTCACCAACCCCTTGTGCGCCTTCCATGCTGCTTCGAGCCATCTGTG 1598
Db 1603 CATGTTCCAGTCCACATCACCTTGTGCGCAATTCATGCGCTCGAGGCCATCTGTG 1662
Oy 1599 TGGCCGATCTATGTTAGTGTCTGTGGAAAGCATTAACCTTGTCTGAAAAAAT 1658
Db 1663 TGGACCAGTTTACGTTAGTGTGTGGAAAGCACAACCTTCAAGTTCCTCAAGAGCTT 1722
Oy 1659 AGTGGTTCCTGATGATGATCCTTTCGAGTGAATGATGATGATGATGATGATGATGATG 1718
Db 1723 CCGTTTGGCTGATGGAGCATTTTGTGTGATGACATGATGATGATGATGATGATGATG 1782
Oy 1719 TTTGTTTGAAGACCCCTTTCATATGAGAAATATGATGATTTGGAATCTCAACAA 1778
Db 1783 TTTGTTTGAAGACCCCTTTCATATGAGAAATATGATGATTTGGAATCTCAACAA 1842
Oy 1779 GTTCACTGAGTGTGTGTGATTCATCTGCAAGAGAGAGATGTTGTGTGAGACAG 1838
Db 1843 AATATACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1902
Oy 1839 CCGCAACCAATGCTTTTCAACATACCAAAAGAGTGAATCCAAACATTAACCCAAAGA 1898
Db 1903 GAGAAACAAAGAGTGCCTTTCACAAATTTTCACAAATGCTTGAAGTCTTCAAGA 1962
Oy 1899 CATAGATGGCAGTGTGAGAAACCCATATCTATTTGAAGCGCTTAAACCTTTGCGCT 1958
Db 1963 CATAGATGGCAGTGTGAGAAACCCATATCTATTTGAAGCGCTTAAACCTTTGCGCT 2022
Oy 1959 TTAAGTCTATCAAGCAAAACCTTATCTCTCAAGCCCTCTCAAGATCTTGACATAC 2018

```

```

Db 2023 ATATTTGTTCAAGACACAAACCTAATTAAGCTCATGAGGCAATCAGAAATTTGAGATTTC 2082
Oy 2019 TCTTGACCCATTTGCAATTCGAGCTCATGCTGTTCACACAGTGAACCAATCATCAAC 2078
Db 2083 ACTTGAAGCATTTACTTTTGAAGCTATTGACAGTGTCTCAGATGATTTGCTGTCAAAA 2142
Oy 2079 TTTCTTACACTTTTGGCCCAATTTGGGCTGTGTAACATGCTTTAACAAGTGAAGCCATCA 2138
Db 2143 GTTAATTCAAATTTGCTCCAAATTTGATTAAGTGAACATGCTTAACATGATGGTGGTCCATCA 2202
Oy 2139 ATCTGTGACATTTGACGATGACCTTAAGCTCAAGTCCAGATTTGCTCAAAAGGCTGTGTA 2198
Db 2203 GTCCATGTGAGTTTGTACAAACACATGATGTGTCAAAATTTGGGCTTGAAGGCTGTG 2262
Oy 2199 GATGCGAGTATTTGATGAGAAAAAACCAAGGCTTGTGCTATGATTTGAGGAGGATTTGG 2258
Db 2263 GATGAAGGTGTTTGTGATGAGAAACCAAGTATGTTTCAAACTAGATGGGATGTTGAAA 2322
Oy 2259 GTTCAAGTATGATCAAGCAAAATGTTGGTGTTCAGTGGCCATGGCAATTTGATTTCTT 2318
Db 2323 ATTTGATTAATG---AGGATTAATGCTGTGAGATGCAAGTCCCTGGCTTGAATTTCAAA 2379
Oy 2319 ATCG 2322
Db 2380 ATTG 2383

```

RESULT 5

AC89523 ID AC89523 standard; DNA; 2498 BP.

AAC89523;

07-MAR-2001 (first entry)

Soybean raffinose synthase coding sequence SEQ ID NO: 23.

Plant promoter; transgenic plant; desired property; ds.

Glycine max.

EP1048733-A2.

02-NOV-2000.

27-APR-2000; 2000EP-0108962.

30-APR-1999; 99JP-0124527.

PR 01-SEP-1999; 99JP-0247211.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI Ishige F, Watanabe E, Oeda K;

DR WPI; 2001-104537/12.

P-PSDB; AAB49400.

New soybean plant promoters useful for generating transgenic plants

with desired properties - Example 2; Page 24-28; 36pp; English.

The present invention provides novel plant promoters which can be used in

the production of transgenic plants which express genes with desired

properties. CC Sequence 2498 BP; 651 A; 521 C; 673 G; 653 T; 0 other;

Query Match 41.5%; Score 976.4; DB 22; Length 2498;
 Best Local Similarity 66.0%; Pred No. 1.3e-303;
 Matches 1507; Conservative 0; Mismatches 741; Indels 36; Gaps 5;

Oy 57 TGACATGTCGTCACCGTTTGAATGACGAGATCGGATTTGACTGTGAAAGCTATTCGTT 116

118 TAATTTGCTTTGTCCATAACCCCTAGAGGATCAAAATTTCTCGCCAAACGACCCCTTT 177
117 TCGTGCATGTTCTCGAGAAACATTTGCTCTCCCTCCCGTAC-----AC 164
178 TCTCACGGAAGTTCCGAAACATAAATAGTCACCCCTTCACCCATGACGCCAAGACTAG 237
165 TTGAGATAGACAGATGCCCGGTTTCGGTTGGTCTTTGTTGATGTGACGCGCTGGAACC 224
238 TAGAACAACAGAGAGACGACACGTCGTAGTGTCTTCGTGGGCTTCACGCGGACGAGCC 297
225 TGAATAGCCAGATGTTGTTTCAATTTGGGAAGCTGAAGATATTCGGTTAAATTTT 284
298 CACAAGCCGACAGCTGGCTTCCCTGGGGAGCTCAGAGAAATTAATTCATGAGCATATT 357
285 CAGTTTAAAGGTTTGGTGGCTACACATGGGTTGGTGGAAATGGGGGATGTTGATC 344
358 CCGGTTTAAAGTTGGTGGACCACTACGTGGGTGGATGCAACGACACGACACTGGAGCA 417
345 GGAGACTCAGATTTGTATCTTGAAGATCAG---ATTGCTGACCGGTATTTTCT 401
418 CGAGACACAGATGATGCTTCTGCAAAAACGACACGCTGGAGCGCCCTTTGTTGAT 477
402 TCTTCGATCGTTGAGGACCGTTCCGACCTGATTCAGCTGGGGATGATCTTTGT 461
478 TCTCCGATCTCCAAAGCCTCTTCGAGCCCTCCGCAACCCGCTTTGGATATTACGT 537
462 CGATGTTTGTGCGAGAGTGTGCTGCGAAAGTTGTTGATGATCGTCCGAATATGTT 521
538 GGAAGCTTTGATGAGAGCGGGGTCGACAGCTGTCTGTGGCTCCAGCTTCGGAGCTGCT 597
522 GATCTTCATGCTGTGATGATCCGTTTGCATGTTTAAAGAGCGATGAATCTGTAG 581
598 ATAGCTCCAGTGTGGCATAGCCGATATCAGTTGCTTAGAGAGCAACTAAATCGTTAG 657
582 GACCACTTGTGAACCTTTCGTTGTTGAGAGAGAACTCCACCGAGTATCTGTGACAA 641
658 GATGCAATTTGGGAGCTTCAAGCTTCTCGAGGAGAAACCGCGCGATGCTCATACAA 717
642 ATTGCTTGTGACAGTGGGACGCTTTTAACTTAACGTTTCAACAGGACGATTAATAG 701
718 GTTGTGTTGTTGATGATGGAGCGCTTTTACTTGAAGGTGACATCCCTCAGGTGTGGGA 777
702 AGGCGTGAAGCATCTGCTGACAGCGGCTTCTCCCGGTTTATGCTTAATGAGCATGG 761
778 AGGGGTGAAGAGGTTGTGAGAGGAGGTGCTCCGAGGATGCTTAATCACAACGG 837
762 TTGGCAATCATGAGACGATTCGATCCCATCAC---CAAGAGGAATGAACCAAC 818
838 GTGGCAAGCATTTTGTACGACGAGAGACCCCATTAACGACCAAGAGGATGTGAAGCAAC 897
819 CGTGGCCGAGCAAAATGCCCTGCTTTTGAATTCGAAGATTAACAAATTCG 878
898 CTCGCGAGGGAGCAAAATGCATGAGGTTGTGAAGTTGAGAGAAATTTACAGTTTCA 957
879 TGACTATGTCATCCCAAGGCCACCGGCCCGGACCGGCGACAGAGGATGAAGCCGTT 938
958 ACAGTATTTAGTGAAGAGATTCG-----AGAAAGTATGGTCTCTT 1002
939 TATAGATGAACCAAGAGATTTAAGACTGTGAGCACTGTTATGTTGGATGCTTT 998
1003 TGTTAGGAGCTTGAAGAAACAGTTTAGAGCTGTGAGAGGTGTATGTGGACCGCT 1062
999 GTGTGATTTGGGCTGCTTCGCGCAGAGTGCCTGGCTTTCGAGCAAGTGTGAT 1058
1063 TTGTGGGTATTTGGGGTGAAGACCAAGGTTCCGGGCAATGCCACAGGCTAAGTTGT 1122
1059 TGAAGCAGTCTTTCAACAGGGCTGCAATGACATGAGAGATTTGGCGTGAATGAAT 1118
1123 CACTCGAAGCTGTCCAAATGACTAAATTTGACAAATGAAGATTTAGCGGTGAAT 1182
1119 TGTTCCTTAAGGTGCGGCTGTCCCGCGGAGAGGCTGAGAGATGTAGCAAGACT 1178

1183 CGTCAGTAACGAGATTGAGACTGTGCTCCACACACACTGGCTCACCTTTTGTAGCAGGGGCT 1242
1179 TCATGCTCATTTGGAAAAAGTTGGGATTCGACGGTGTAAAGATTGAGATTATCCACTATT 1238
1243 CCACTCCGTTTGGATCTCGGGGTATTGACGCTGTTAAGGTTTACGTTTACACTTGTCT 1302
1239 GGAAGATTGTTGTGAAGACTATGAGAGGAGAGGATTTGGCAAGGACATATTACAAAG 1298
1303 CGAAGTGTCTTCGAGGAATAGGTGGCGCTGTGAGCTAGCCAAAGGCTATTACAAAG 1362
1299 AATGACCAATCAATTAATTAACATTTTAAAGSAATGAGTATTGCAATGTGAACA 1358
1363 GCTACATGCTTCGTTGAAGAGATTTCAAAAGCAATGGGCTATTGCGGACATGAGAGA 1422
1359 TTGTAAGCATTCATGCTTCCTTGGCAGGAGGATCTGCTGCTGGTGGTGTGATGA 1418
1423 TTTTAATGACTTTTCTCTCTTGTACCGAAGCCATAGCCCTTGGGCGCTAGAGATGA 1482
1419 CTTTGTGCGACGACCCCTCTGCTGATCCAAACGCTAGCTTTTGGCTCCAAAGATGCA 1478
1483 TTTTGTGCTAGTATCCCTGTGAGATCCAAATGACAGTATTTGGCTCCAAAGGTGTCA 1542
1479 CATGTTCAATTTGCCAAGACAGCTTTGTGATGGGCACTTATCCACCTGACTGGGA 1538
1543 CATGGTCACTGTGCTTACAAACAGCTTGTGATGGGAAATTTTATTCAGCGGATTTGGA 1602
1539 TATGTTCAATCCACCAACCCCTGTGCGGCTTCATGCTGCTTCGAGCCATCTCTGG 1598
1603 CATGTTCCAGTCCACTCACCTCTTGTCCGAATTCATGACGCTCTTAGGSCATCTCTGG 1662
1599 TGGCCGATCTATGTTAGTATCTGTGGGAAGCACTTAATCTTGTATCTTCTGAAAAACT 1658
1663 TGGACAGTTTACGTTAGTATGTTGTGGAAGCAACTTCATGCTCACAAGGCT 1722
1659 AGTGTCTCCGATGATGATCACTCTGAGTGTGATGCTATGCACTCCGACCTGGCATG 1718
1723 CGCTTTGCTGATGGACGATTTTGTGCTTCAACCTATGCACTCCCAACAGAGATG 1782
1719 TTTGTTTGAAGACCTTTGATATGAGAAACTATGATTTGAATTTGAATCTCAACAA 1778
1783 TTTGTTTGAAGACCTTGTGATGATGAGAAACATGCTCAAAATTTGGATCTCAACAA 1842
1779 GTTCACTGAGTGTGTTGCTATTCACATGCTCCAGAGAGAGATGCTGTGAGACAG 1838
1843 ATATACAGGTTGTTGGTCTATTTAATTGCAAGAGGTTGGTGTGTCGCTAATCTAG 1902
1839 CCGCAACCAATGCTTTTCAATCAATCAAAAGAGGATGATCCAAACTAACCCAAAGA 1898
1903 GAGAAACAAAGTGTCTGATTTTTCACAACTGTGACATGCTTAGCGATCTTACAGA 1962
1899 CATGATGAGCAGTGAAGAAAAACCTATCTCTATTGAAGGCTTAAACCTTTGCGCT 1958
1963 CATTAATGAGAGCAATGGGAAAGCCCAATATGCAATAAAGGATGAATGTGTCTGT 2022
1959 TTAATCTTATCAAGCAAAACTTATCTCTCAAGCCCTCTCAAGATCTTGACTAGC 2018
2023 ATATTTGTTCAAGGACCAAACTAAAGCTCATGAAGCATAGAGAAATTTGAAATTTT 2082
2019 TCTTGACCATTCGATTTGAGGCTCATCACTGTTTACCAAGGACCAAACTATCCAAAC 2078
2083 ACTTGAGCCATTTACTTTTGAAGTATTTGACAGTGTCTCAAGTATTTGTCTGTCAAAAA 2142
2079 TTCTTACACTTTTGGCCCAATTTGGGCTGTGGAACATGCTTAACTAGTGAAGCATCA 2138
2143 GTTAATTAATTTTGTCCCAATTTGATTTAGTAACATGCTTAAACATGCTGTGAGCATTT 2202
2139 ATCTGTGACTATGAGATGACCTTAAGCTCAGATTTGTTGTTCAAGGCTGTGCTGA 2198
2203 GTCCATGAGGATTTGACAAACCATAGATGTGTCAAAATTTGGGGTTAGGGGTTGTGGGA 2262
2199 GATGAGATTTTGAATGCAAAAAACCAAGGCTGTCTATTTAGTGGGAGGATGTTGG 2258
2263 GATGAAGGTTTGTGATCAGAGAAACCAAGTTAGTTGCAACTAGATGGGTTAGTTGTAAA 2322

OY	1188	TTTGGAAAAATTGGGATTCACCGCTGTAAAGATTGGACGTTATCCACCTATGAGAACTT	1247
Db	1423	TTTGGAAAAATTGGGAAATTGATGAGTCAAAAGTTGATGTCATCCATTTGTTGGAGATAT	1482
OY	1248	GTTGTAAGACTATGAGGAGAGTGTGATTTGGCAAAAGCATATTACAAAGCAATGACCA	1307
Db	1483	GCGAGAGACTATGAGAGAGATTTGAACTACCAAAACATCTATTAAGCCAAATTAACGA	1542
OY	1308	ATCAATAAATTAACATTTTAAAGAAATGAGCATCTGCAAGTATGAAACATTGTAACGA	1387
Db	1543	ATCAGTGGTAGGATTTCAAAGCAACGGTGTGATTTGTACATTGAGAGCACTGCAACGA	1602
OY	1368	CTTATGTTCTTCCTGGCAACGAACTATCTCTCTGGTCCGTTGGTGAATGACTTTGGTG	1427
Db	1603	TTTTCATGCTCTTGTGACTAGAACATTTGTCTGTGCTCGTTGGGATGACTTTTGGCC	1682
OY	1428	CACGAGCCCCCTGTGTATCCAAACGCTACGTTTGGCTCCAGATGTCACATGTTTCA	1487
Db	1663	AACGATACCGCTGGAGATTAATATGGTACATATTTGGCTCCAAAGGCTGTCAATGTGCA	1722
OY	1488	TTTGGCAACAGCACTTTGGATGGGAACTTATCCACCCCTGAGCTGGGATATTTGCA	1547
Db	1723	TTTGGTCCACAAATGCTTATGATATGGGAAACTTATACACCTGACTGGGACATTTCCA	1782
OY	1548	ATCCACCCACCCTTTGCGCCTTCCATGCTGCTCTGAGCCATCTGTGGCCGAT	1607
Db	1783	ATTCACACACCCTTGTGCTAAATTTATGCTGTGATCTCTGCGATTTCTGGGACCAAT	1842
OY	1608	CTATGTTAGTATCTTTGGGAAAGCATTAACCTTGTATCTTGGAAAAAATAAGTCTGCC	1667
Db	1843	TTATGTTAGTATTTTGTGGCAAGCATACATCCCTTGTCTCAAAAGGCTGTCTTGGC	1902
OY	1668	TGATGATTCGATCTTGGAAAGTAGTACTATGACACTCCGACGCGCATTTGTTGTTGA	1727
Db	1903	TGATGTTGCGATCTTCTGTGGCAGATACCATGACTCTTACTAAGATTTGCCATTGTT	1962
OY	1728	AGACCCTTGGATTAATGAGAAACATTTGCTTAAGTTTGGATCTTCACAAAGTTACATGG	1787
Db	1963	AGATCTCTTGGACATGTGCAAAACATGCTCAAAATTTGGAACTCCAAACAGTACAAATGG	2022
OY	1788	AGTATTTGGTCATTTCAACTGCCCAAGAGAGAGATGGTCTGTGAGAACCCCGCAACA	1847
Db	2023	AGTCTTGGAGTCTTCAATTTGCCCAAGAGAGAGGTGAGACCTGTAGTCTCGAAAAATCT	2082
OY	1848	ATGCTTTTCACATATCTCAAAAGAGATGACATTCCAAACCTAACCCAAAAAGACATGAA	1907
Db	2083	ATGTTTCTCAGAGATTTCAAAACCTATTTCTGTGCAAGCAACATGCCAAAAAGATGTTGAA	2142
OY	1908	GCAAGTGGAGAAAAACCCATCTCTATTGAAGCGCTTAACCTTTGGCGTTAACTCTTA	1967
Db	2143	GGAACAGGAAACAAAGCCATTTCCCATCAAAAGAGTGAATTTTGGCATTTACTTCCAC	2202
OY	1968	TCAAAGCCAAAAAATTAATCTCTCTTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACC	2027
Db	2203	CAAGAAAAAAAGATAATCTCTTCCACAACTATCTGACACATCTGAATATCACTTGATACC	2262
OY	2028	ATTGGAATTTGAGCTCTACCTGTTTCACAGAGCAAACTCATTCACAACCTCTCTACGA	2087
Db	2263	CTTGATTTAGAGATTATTTGATCTCTCTCCGATGACAAATTTTACCCTTGGAGTGCATGCC	2322
OY	2088	CTTTTGGCCCAATTTGGCTGTGTGAACATGCTTAACACTAGTGGAGCCATTCAAATCTGTGGA	2147
Db	2323	ATTTGCACCCATAGGATTATTAACATGTCTCAACCGCGGAGGGGAGTCAAGTCTTTGGA	2382
OY	2148	C-----TATGAGCATGACCTAAGCTCACTCGAGATTTGGTGTCAAAAGGCTGTGGAGAT	2201
Db	2383	CATCAGTAGAGATTAATGAGATTAAGATGTTCAAGTTGGTATTTAAAGGGGCGGAGAAAT	2442
OY	2202	GCGAGTATTTCACGAAAAAACCAAGGCGCTGTGCTATATGATGGGAGAGATTTGGGTT	2261
Db	2443	GATGCTTTATTCATCAGAAAAAGCCAAACCGCTGTAGAGTTAATGAGAAAGCATGTGAGTT	2502
OY	2262	CAAGTATGATAGGACCAAAATGTTGTT	2288

[illegible]

OY 402 TCTTCCGATCGTTGAGGACCGCTCCGAACCTCGATTGACGCTGGGATGATGACTTGT 461
DB 417 TCTCCCGATCTCCAAAGCCTGCTCCGAGCCCTCCGCAACCGGTTGGATGATTACGT 476
OY 462 CGATTTTGTCTCGAGAGTGGTTCGTAAGTGTGTATGATGATGCTTCCGAATATGT 521
DB 477 GGACGTTTGGATGAGAGCGGCTGCACAGTGTCTCTGCTCCACCTTGGGAGCTGCT 536
OY 522 GTATCTTCACTGGTGTGATGATCGCTTTCACCTGTTTAAAGAGCGGATGAAGATCGTAG 581
DB 537 ATACGTCACAGCTTGGCCATGACCCGATACGTTGCTTAGAGAGCAACATGAAGTGGTAG 596
OY 562 GACCCATCTTGGAACTTTTCCGCTTGTGAGAGAGAAAGACTCCACAGATATCGTGACA 641
DB 597 GATGATTTGGGAGCTTTCAGACCTTCTCGAGAGAAACCGGCGCAGATGATGTAGACAA 656
OY 642 ATTGCGTGTGTCACGTCGAGCGCTTTTACCTAACGGTTCATCCACAGGCGTAAATGA 701
DB 657 GTTGTGTTGTTTACATGGGACGCTTTTAACTTGAAGGTGATCCCTCAGGTGTGGGA 716
OY 702 AGCGGTGAGGACATCTGTCGACGGGCTTGTCTCCGCTTAACTGCTAATGAGAGATGG 761
DB 717 AGGGGTGAAGGGCTTGTGAGGAGGGTGCCTCCAGGATGCTTATGACGACGG 776
OY 762 TTGGCAATCCATCGGACACAGATTCGATCCATCAC--CAAGAAGAATGACCAAC 818
DB 777 GTGGAGAGCATTTGTGCACGACGAGGAGCCCATTAAGCAAGAGGATATGAGAGAAC 836
OY 819 CCGTCCCGGCGAGCAATGCCCTGCCGCTTTTGAATTTCCAAAGAAATTAACAAATTCGG 878
DB 837 CTCGCCAGGGGAGCAATGCCATGAGTGTGTAAGTTGAGAAATTAACAGTTCAG 896
OY 879 TGACTACGTCATCCCAAGGCCACCGGCCCGGAGCGGCGCAAGAGGGGATGAAGCGCT 938
DB 897 ACAGTATTTGTAAGGAAAGGATTTCTG-----AGAAGGATATGGGTGCTT 941
OY 939 TATAGATGAATCAAGAGAGAGATTTAAGACTGTGAGCATGTTTATGTTTGGCATGCTTT 998
DB 942 TGTTAGGACCTTGAAGGAACAGTTTAGAGCGTGGAGCAGTGTATGT--GCAGCGCT 998
OY 999 GTGTGATATTTGGGCTGGCTTCCGCCGAGAGTGCCTGGCTTCCAGAGCACGTGTAT 1058
DB 999 TGTGTGGAATTTGGGCTGGGCTCAGACCCAAAGTTCCGGGATGCCGCCAAGCTTAAGTGT 1058
OY 1059 TCAGCCAGTGTCTTCCACGAGGCTGCAGATGACGATGAGGATTTGGCGGTGATTAAGT 1118
DB 1059 CACTCCGAAGCTGTCCATGACTAAATTTGACAAATGAAGATTTAGCGGTGATTAAGT 1118
OY 1119 TGTCTTATTAAGTTCGGGCTGCTCCGCCGAGAGAGGCTGAGAGATGTACGAAGACT 1178
DB 1119 CGTCAATGAAGAGTGTGAGCTGTGCACACACCTGCTACCTTTGTACGAGGGGT 1178
OY 1179 TCATGCTATTTGGAAAAAGTTGGGATCGACGCTGTTAAGATTTAGCTTATCCACTATT 1238
DB 1179 CCACCTCCGTTTGAATCTGCGGGATTTGACCGGTATTAAGTTACGTTATACATGCT 1238
OY 1239 GGAGATGTTGTGAAGACTATGAGGAGGAGTGTGATTTGGCAAGGCAATTTACAAAGC 1298
DB 1239 CGAGATGCTATCCGAGCAATACGCTGGCGGTGTTGAGCTAGCCAAAGCTTATTACAAAGC 1298
OY 1299 AATGACCAATGAATAAATTAATTTAAAGAAATGAGTCAATGCAATGAGAAACA 1358
DB 1299 GCTCACTGCTGGTGAAGAACATTTCAAGGCAATGCGGTCATTTGCGAAGCATGAGACA 1358
OY 1359 TTGTAAACGACTTATGTTCTTGGCAGCAAGCTATCTCTTGGTCTGTTGGTGATGA 1418
DB 1359 TTGTAAATGACTTCTTCTTCTTGGTGAACCAAGCCATTAAGCCCTTGGCGGTGAGATGA 1418
OY 1419 CTTTGTGTCAGGAGCCCTGTGTGATTCAAAGGATGATTTTGGCTCAGAGATGTCA 1478
DB 1419 TTTTGTGTCAGTGAATCCCTGTGAGATTCAAATGACGATTTGGCTCCAAGGGTGTCA 1478

OY 1479 CATGTTTATTTGTGCCAACGACAGCTTGTGATGGGGAATTTCAATCCACCTGACTGGGA 1538
DB 1479 CATGGAGCATGCTGCTACAAACAGCTTGTGATGGGGAATTTTAACTCCGAGATTGGGA 1538
OY 1539 TATGTTTCAATCCACCAACCCCTGTGGCCCTTCCATGCTGCTCTCGAGCCATCTCTGG 1598
DB 1539 CATGTTTCAATCCACCAACCCCTGTGGCCCTTCCATGCTGCTCTCGAGCCATCTCTGG 1598
OY 1599 TGGCCGATCTTATTTAGATTTCTGGGAAAGCAATTAATGATCTTCTGAAAAAATCT 1658
DB 1599 TGGACCAATTTACGTTAGATGTTGTTGTGGAAGCAACAACTTCAAGTCTCAAGAGCCT 1658
OY 1659 AGTGTCTCTGATGATGATGATCTTTCGAAGTGAATGATGACATCCGACTCCGATGCGATTG 1718
DB 1659 CCGTTGCTGATGAGAGACTTTTGGAGACTTTTGTCTGACACATGACATCCCAACAGAGACTG 1718
OY 1719 TTTGTTTGAAGCCCTTTCGATTAATGAGAAACTATGCTTAAGATTTTGAATCTCAACA 1778
DB 1719 TTTGTTTGAAGCCCTTTCGATGATGAGAAAGCAATGCTCAAAATTTGGAATCTCAACA 1778
OY 1779 GTTCACTGAGATTTGGTGTGATTCACATGCGCAAGAGAGGATGGTGTGAGAGACAG 1838
DB 1779 ATATACAGGTGTTTGGTGTGATTTTAATTTGCCAAGAGAGTGGTGTGCTCCGTAATGAG 1838
OY 1839 CCGCAACCAATGCTTTTCAATATCTCAAAACGAGTACATCCAAACCTAACCACAAAGA 1898
DB 1839 GAGAAACCAAGAGTGGCTGTGATTTTACAAACCTGTGACATGCTTAGCAGATCTCAGAGA 1898
OY 1899 CATGATATGACAGTGGAGAAACCCCTATCTCTATTGAAGCGCTTAAACCTTTGGCT 1958
DB 1899 CATGATATGAGCAATGGGAAAGCCCAATATGCAATTAAGAGGATGAATGTGTTGCTGT 1958
OY 1959 TTACCTATCAAGCCAAAAATATCTCTCAAGCCCTCAAGATCTTGACATGAC 2018
DB 1959 ATATTTGTTCAAGGACCAACCAATTAAGCTCATGAAGCATAGACAAATTTGGAATTTTC 2018
OY 2019 TCTTGACCCATTCGAATTCGAGCTCATCTGTTTCCACGATGAGCAAACTCATCCAAAC 2078
DB 2019 ACTTGAGCATTTACTTTTGTGACTATTTGACAGTGTCTCAGATGATGTCTGTCAAAAAA 2078
OY 2079 TTCTGTACATTTGGCCCAATTTGGGCTGTGTAACATGCTTAACACTATGAGGCCATCCA 2138
DB 2079 GTTAATTCATTTTGTCTCAATTTGATTAAGTGAACATGCTTAACACTGAGTGTGCTTCA 2138
OY 2139 ATCTGTGACATGATGAGATGAGCTAAGCTCAGTCGAGTATTTGATGGGAGATCTTGG 2198
DB 2139 GTCCATGAGTTTGAACACCAATGATGATGTGTCAAAATTTGGGCTTAGGGGCTTGTGGGA 2198
OY 2199 GATCGAGTATTTGATCGAAAAAACCAAGGCTTGTCTATTTGATGGGAGATCTTGG 2258
DB 2199 GATGAAGGTTTGTGATCAGAGAAACCACTTAGTTGCAAAATGATGGGCTTAGGTGTAA 2258
OY 2259 GTTCAAGTATGATCAGGACCAATGCTGTTCAGATGCCATGSCCAATGATTTCTTC 2318
DB 2259 ATTTGATTAATG--AGGATTAATGCTGAGAGTGCAGAGTTTCCCTGCTGAGTTCAAA 2315
OY 2319 ATCG 2322
DB 2319 ATTTG 2319

RESULT 8
AAZ10002
ID AAZ10002 standard; cDNA to mRNA; 2497 BP.
XX
XX AAZ10002;
AC
XX
XX 26-OCT-1999 (first entry)
DT
XX
XX Nucleic acid encoding a raffinose synthase protein.
DE
XX
XX
KM Raffinose synthase; plant; sucrose; raffinose; ss.
XX

Db	1722	CGCTTGGCCATGATGGAGACGATTTTGGGTTGTCAACACATATGACATGCCACACAGACTG	1781
Qy	1719	TTTGTTGAAGACCCCTTTCATATATGAGAGAAACATATGCTTTAAGATTTGGATCTCAACAA	1778
Db	1782	TTTGTGTGAAGACCCCTTGCATGATGGGAAAGCAATGCTCAAAATTTGGAAATCTCAACAA	1841
Qy	1779	GTTCACTGAGATGATTTGTGCATTTCAACACGCCAAGAGAGAGATGCTGTCTGAGACACG	1838
Db	1842	ATATATCAGGCTTTTGGGCTATTTTAAATTCACAAAGAGAGGTGGGTGGTCCGTAACTAG	1901
Qy	1839	CCGGAACCATATGCTTTTTCACATATCTCAACAAAGAGATGATCTCAAAACCTTAACCCAAAGA	1898
Db	1902	GAGAAACAAAGAGTCCCTCTGTAATTTTTCACAAACTGTGACATCTTTAGCAGTCTCAACGA	1961
Qy	1899	CATGAAATGGCAGACAGTGGAGAAAAACCTATCTCTATTTGAAGCGCTTAAACCTTTTGCCT	1958
Db	1962	CATTGAATGGAGCAATGGGAAAGCCCATATGATCAATAAAGGATGAAATGTGTGCTGT	2021
Qy	1959	TTACCTCTATCAAGCCAAAAAATTTATCTCTTCACAAAGCCCTCTCAAGATCTTGACATAGC	2018
Db	2022	ATATTTGTTCAAGGACCAACAACTAAAGCTCATGAAGGCATCAGAAATTTGGAAGTTTC	2081
Qy	2019	TCTTGACCATTTGGAAATTTGAGAGCTCATCTGTTTACACAGTACCAACATCATCAAC	2078
Db	2082	ACTTGACCATTTTACTTTTGTAGCTATTGACAGTGTCTCCAGTGAATTTGTCTCTCAAAAAA	2141
Qy	2079	TTCCTCACTTTGCCCAATTTGGGCTGTGTGAACATGCTTAACACTAGTGGAGCCATCCA	2138
Db	2142	GTTAAATTCATTTGCTGCTCATATGGAATTAAGAACATGCTTAACACTGGTGGTCCATCCA	2201
Qy	2139	ATCTGTGACTATGACAGATGACTTAAGCTCAGTCGAGATTTGTTGCATAAAGGTGTGTGCA	2198
Db	2202	GTCCTATGGAGTTTGACACCAACCATATGATGTGCAAAATTTGGGGTTAGGGGTTGTGGGGA	2261
Qy	2199	GATCGAGTATTTGGCATGCAAAAAAACCAAGGGCTTGCTATTGATGGGAGAGACTGG	2258
Db	2262	GATGAAGGTGTTTGGCATGACAGAAACCACTGATGTTGCAAACTAGATGGGTAAGTTGTAAA	2321
Qy	2259	GTTCAAGTATGATCAGAGCAAAATGTTGGTGTTCAGATGCCATGGCCATATGATCTTTC	2318
Db	2322	ATTGATTTATG---AGGATTAATAATGCTGAGAGTGCAGATTTCCCTGGCTAGTCTTCAAA	2378
Qy	2319	ATCG 2322	
Db	2379	ATTG 2382	
RESULT 9			
AAV40800			
ID	AAV40800 standard; cDNA to mRNA; 2746 BP.		
XX	AAV40800;		
AC			
XX	23-SEP-1998 (first entry)		
DT			
XX	Broad bean raffinose synthetase coding sequence.		
DE			
XX	Raffinose synthetase; metabolism modification; food additive;		
KW	gastrointestinal flora; broad bean; ss.		
KW			
XX	Vicia faba.		
OS			
XX			
FH	Key		
FT	CDS		
FT			
XX	EP849359-A2.		
PN	24-JUN-1998.		
PD			
XX	18-DEC-1997; 97EP-0122417.		
PF			

PR	18-DEC-1996;	96JP-0338673.
XX		
PA	(SUMO) SUMITOMO CHEM CO LTD.	
XX		
PI	Oeda K, Wantanabe E;	
XX		
DR	WPI; 1998-324670/29.	
DR	P-PSDB; AAM57886.	
XX		
PT	New nucleic acid molecule encoding plant raffinose synthetase -	
PT	capable of producing raffinose, used as food additives with	
PT	beneficial effects on gastrointestinal flora	
XX		
PS	Claim 1; Page 26-29; 44pp; English.	
XX		
CC	This sequence encodes the broad bean raffinose synthetase of the	
CC	invention. The raffinose synthetase is capable of producing raffinose by	
CC	combining a D-galactosyl group through an alpha (1-6) bond with a	
CC	hydroxyl group attached to the carbon atom at position 6 of a D-glucose	
CC	residue in a sucrose molecule. The DNA can be used to modify metabolism	
CC	of a host organism by introducing into the host organism or its cell so	
CC	that the content of the raffinose family oligosaccharides in the host	
CC	organism or cell is changed. Raffinose oligosaccharides are useful as	
CC	food additives with beneficial effects on the gastrointestinal flora.	
XX		
XX		
Sequence	2746 BP; 780 A; 552 C; 621 G; 793 T; 0 other;	

Query Match	40.7%	Score 956.6	DB 19	Length 2746
Best Local Similarity	66.6%	Pred. No. 3,5e-297		
Matches 1421	Conservative 0	Mismatches 694	Indels 18	Gaps 3
QY 193	GTTTCCTTTGTTGGATGACCGCTCGAGACCTCGATAGCCGACATGTTTTCGATTGGG	252		
Db 359	GGTTGTTTCGTTGGTTTCACTCCACCGAACCCMAAGCCACACGTAATTCCACTCCGC	418		
QY 253	AAGCTGAAGGATATATTCGGTTTATGAATATTTTCAGGTTTAAAGTTTGGTGACTACAC	312		
Db 419	AAACTAAAGGAAATCAATTCATGAGCATATTCGGTTCAAGTTTGGTGACAACTCAC	478		
QY 313	TGGGTTGGTCCAAATATGATGGGGAATCTTGAAATGGAGACTCAATTTGATCTTTGAGAG	372		
Db 479	TGGGTCGGAGAACAAAGGACAGGAACCTACAAACGAAACCAATATTTAATCTGGACAAA	538		
QY 373	TCAGATATTC---TGGTCGACCGGATATGTTTCCCTTCCTCCGATGGTGAGGACGGTTACGA	429		
Db 539	AACGACATCCCTTCGGACGACGACCTATGTCTTACTCTCCCAATCTTGAAMACACTTTCCCA	598		
QY 430	ACCTCGATTTCAGCCCTGGGGATGATGACTTTTGTGATGTTTGTGTGAGATGGTTCGTG	489		
Db 599	ACCTCAGTCCAACCCGGGTCTCAACGATCATATAGGATGTCGGTGAAMAGCGGTTCAACA	658		
QY 490	AAAGTTGTTGATGCATCTGTTCCGAAGTATGTTGATCTTCAATGCTGTGATGATCCGTTT	549		
Db 659	CATGTCACCCGGGTCAAGCTTCAAACCATGTCTTACATCCATCTCAGTAACGCCCATATAC	718		
QY 550	GCACCTTGTAAAGAGCGCGATGAAAGATTCGAGAGCCCATCTTGGAACTTTTCCGTTGGTG	609		
Db 719	AGTATACTTAAAGAGACGAGTTAAAGTAACTCAAACTCACTTAGGACATTTCAAGACTCTT	778		
QY 610	GAGGAGAAAGACTCCACGAGTATCGTGGCAAAATTCGTTGGTGCACTGGGACCGGTTT	669		
Db 779	GAAAGAAAAACGACGACCTAGTATTTATAGCAAAATTCGGTTGGTGACGCTGGGATGCTTTT	838		
QY 670	TACCTAAGCGTTTCATCCACAGGCGGTAAATAGAAGCGGTAGGCAATCTCGTCCACGCGGT	729		
Db 839	TACTTGAAGGTTCATCCAAAAGGTATATGGGAAGGTGTAAGTCTCTCTCACAGATGGTGGT	898		
QY 730	TGTCTCCCGGTTTGTGTCCTATTCGACGATGGTGTTGGCAATTCATGAGACATGTCGGAT	789		
Db 899	TGTCTTCCCGGTTTGTCTATATTCGACGACGCGTTGGCAATTCATTTTGTATGACGATGAC	958		
QY 790	CCCATCCACCAAGGAAGATGAACCAAAACCGTTCGCGGCGAGACAAATGCCCCTCTTT	849		

D	b	959	GATGAGAATGATTAGAGATATGAATGAACCCAGACTCTACGCCGGGGAACCAATGCTATGCACGACTT	1018
O	y	850	TTGAATTCACAAGAAATATCAAAATTCGGTAGCTACGTCAATCCAGAGCCACCGGCCCC	909
D	b	1019	GTAATAATACGAAGAGAAATCTAAGATTAGAGAAATATGAATAATCTGA-----A	1066
O	y	910	CGACCCGCCCAAGGGGAGTGAAGCGCTTATATGATGAACCTCAAGGAGAGCTTTAAAGCT	969
D	b	1067	AATGAGGGAAGAAAGGTTGGGGTGGTTTGTACGGATTTTAAAGAAAGAGTTGGGAGT	1126
O	y	970	GTGGAGCATGTTTATGTTTGGACATGCTTTGTGGATATTTGGGGTGGCCCTTGCGCCGAG	1029
D	b	1127	GTGGAGAGTGTATTATGTTTGGCATCGCCCTTTGGGGTATTTGGGGCGGTTTAGGCTCGA	1186
O	y	1030	GTGCTGTGCTTGCCTGTAGGACAGTGTGATTCAGCCAGTCTTTCACAGGCTGCAGATG	1089
D	b	1187	GTGCATGGGATGCCGAAGAGTAGGGTTGTGTCCGAAGGTGTCTCAGGGGTTTGAAGATG	1246
O	y	1090	ACGATGGAGATTTTGGGGGTGGATTAAGATTGTCTTCATTAAGTGGGGTGGTCCCCCG	1149
D	b	1247	ACGATGGAGATTTTGGGGGTGGATTAAGATTGTGTAGAACGGTGTGGGGCTAGTCCGGCA	1306
O	y	1150	GAGAAAGCTGTGAGATGTACGAAGAGCTTCATGCTCATTTTGGAAAAAGTTGGATGTAC	1209
D	b	1307	GATTTTGCACATGAGATGTTTGATGGGCTTCACTCTCATTTTGGAGTGGCGGGAATTAC	1366
O	y	1210	GGGTTTAAGATTGACGTTTATCCACTATTGGAGATGTTGTGTGAAGACTATGAGAGGAGA	1269
D	b	1367	GGTGTAAAGTTGATGTATTATCCATCTGCTTGAGTTACTATCAGAGGAATATGTGGACGA	1426
O	y	1270	GTGCATTTGGCAAGAGCATATACAAAGCANTGACCAATTAATTAACATTTTAA	1329
D	b	1427	GTTGGAGTCAGCAAGAGCTTATACAAAGCACTAACCTCATGATAGAAACATTTCAAA	1486
O	y	1330	GGAATGGAGATCTGTGACAGATGAAGAAACATGTAACAGTTCATGTTCTTGGCAGCGAA	1389
D	b	1487	GGCATGTGTATTTGCTATGACATGAGACATGTGCAAGACATCTTCTCTCCGCAACGAA	1546
O	y	1390	GCTATCTCTTGTGCTGCTGTTGGTGTATGATCTTTTGGTGCACGAGACCCCTGTGGTATCA	1449
D	b	1547	GCCATATCCCTCGGCGCGCGTGGAGATATTTTGGTGTCTGTATCATCATCTGTGATCCA	1606
O	y	1450	AAGGTTACGTTTGGCTTCCAAAGATGTACATGTTTCATTGTGCCAAGCAGCTTGTGG	1509
D	b	1607	AATGATATATTTGGCTCCAAAGGTGTGCATGTGCATTTGTCCCTACACAGTTTATGG	1666
O	y	1510	ATGGGAACTCATGCACACCCGTGAGTGGGATGTTCATCATACACCACCCCTGTGCCGC	1569
D	b	1667	ATGGGAATTTTCATTCAGCCAGATTTGGACATGTTTCAGTCCACTCATCTCTGTGCTGAA	1786
O	y	1570	TTTCATGCTGCCCTCTCGAGCCATCTGTGTGGGCCGATCTATGTATGATCTCTTGGA	1629
D	b	1727	TTTCATGCGCGCTCAGAGGCAATATCCGGGGGCAATTTATGTTATGATATGTGTGTG	1786
O	y	1630	AAGCATACTTTGATCTTTGAAAAAACTATGTGCTTCTGTATGATCGATCTCTTGAAGT	1869
D	b	1787	AATCACATTTTCAAGTGTCTCAAAATCTCTGTGTTTCCGAGGTCTTATCTTGGCTGT	1846
O	y	1690	GAGTACATGACACGCCGAGCTCGGAGATGTTGTTTGAAGACCCCTTGATTAATGGAATA	1749
D	b	1847	CACATTTACGACCTCCCTACAAAGATGTGTTTGAAGACCCCTTGATTAATGGAATA	1906
O	y	1750	ACTATGCTTAAGATTTTGGATCTCAACAAGTTTACTGAGAGTATTTGGATCTTCAACATGC	1809
D	b	1907	ACAATGCTGAAAAATTTTGGAAATCTCAACAAATATACAGGTGTTTGGGCTTTTCAACTGC	1966
O	y	1810	CAAGAGAGAGATGTTGTCGTAGAGACGCGCCAACCAATGCTTTTCAACATCTCAAAA	1869
D	b	1967	CAAGGTGGTGGTGTCTCTGAGGACCGCGCAAAACAAGTGTATTCGAATTTTCAACG	2026
O	y	1870	CGAGTGCATTCAAAACTAACCCAAAAAGCATATGATGCGACAGTGTGAGAAAAACCTATC	1929
D	b	2027	CGGGTGCATCTTATTCGACATCCCGGAAGCATTTGAATGTTGCAATGGGAAAAATCTCAATG	2086

QY	1930	TCATATGAAAGCGTTAAACCTTTGGCGCTTACCCTATATCAAGCCAAAAAACTTATTCCTC	1989
DB	2087	AGCACCAAGGCTGGGATTTTTTGTCTGTATTTTTTCAAGAGAGAAATTAAGGCTC	2146
QY	1990	TCACAGCCCTCCAGATCTTGACATAGCTCTGACCCATTCGAATTCGAGCTCATCACT	2049
DB	2147	ATGAAGTCTTCTGATACATTTGAAAGTTGGCTTGACCATTAGTTTASGCTATATGACA	2206
QY	2050	GTTCACCAAGTGACCAAACTCATCCAAACTTCTATACACTTTGCCCAATTGGGCTGGTG	2109
DB	2207	GTGCTCCACGAGAAAGCTGTTTTCGAAAAGGTTATACATTTTGACCCAGATTGGTTAGTG	2266
QY	2110	AACATGCTTAACACTATGTGGAGCCATCCATCTGAGACATATGAGATATGACCTAATGCTCA	2169
DB	2267	AACATGCTGAACCTGTGGTGGTGCGATTCACTCTCGAGATTGATGATATGATGCAAGTTTG	2326
QY	2230	GCTTGCTGATATTGATGGGAGAGATGTTGGTTCAGTATGATCAGAGACCAATGCTGGTG	2289
DB	2387	TGCTGTCAAAATTTGATGGCGTTAAGGTGAATTTCTTTATG---AGACAAAAATTCGCAAGA	2443
QY	2290	GTTCAGTGCCATGGCCAAATTTGATTCCTTCATCG	2322
DB	2444	GTTCAATTCTGTGGCCTAGTCTTCAACATTTG	2476
RESULT 10			
AAZ10001			
ID	AAZ10001	standard; cDNA to mRNA, 2746 BP.	
XX	AAZ10001;		
AC			
XX			
DT	26-OCR-1999	(first entry)	
XX			
DE		Nucleic acid encoding a raffinose synthase protein.	
XX			
XX		Raffinose synthase; plant; broad bean; sucrose; raffinose; ss.	
KM			
XX			
OS		Vicia faba.	
XX			
FT	Key	Location/Qualifiers	
FT	CDS	101..2500	
FT		/*tag= a	
FT		/product= "raffinose synthase"	
XX			
PN	JPI1215984-A.		
XX			
PD	10-AUG-1999.		
XX			
PF	12-DEC-1997;	97JP-0342899.	
XX			
PR	28-NOV-1997;	97JP-0329006.	
PR	18-DEC-1996;	96JP-0338673.	
XX			
PA	(SUMO) SUMITOMO CHEM CO LTD.		
XX			
DR	WPI: 1999-511112/43.		
DR	P-PSDB; AAY30142.		
XX			
PT	New raffinose synthase gene -	is prepared from a plant material	
XX			
PS	Claim 6; Page 21-24;	40pp; Japanese.	
XX			
CC	The present sequence encodes a raffinose synthase protein. The		
CC	sequence is isolated from plant material of broad beans. The		
CC	protein forms raffinose by complexing alpha(1 to 6)- D-galactosyl		
CC	hydroxyl group of the 6C of D-glucose residue in sucrose molecules.		
XX			
SQ	Sequence 2746 BP; 780 A; 552 C; 621 G; 793 T; 0 other;		

Query Match	40.7%;	Score 956.6;	DB 20;	Length 2746;
Best Local Similarity	66.6%;	Pred. No. 3.5e-297;		
Matches 1421; Conservative	0;	Mismatches 694;	Indels 18;	Gaps 3

Oy	193	GGTTGCTTTGTGGATATTCGACGCGGTGGAAACCGMATAAGCCGACATGTTTTCGATTGGG	252
Db	359	GGTTGTTTCTGTTGGTTTCAACTCACCAGAACCCAAAGCCACCACTGATGTTTCCACTGGCC	418
Oy	253	AAGCTGAGGATATTCGGTTTATGACTATTTTCAGGTTTAAAGTTTGGTGACTACACAC	312
Db	419	AAATCAAAAGGAATCAAAATTCATGACATATTCGGTTTCAAAAGTTTGGTGACAACATCAC	478
Oy	313	TGGGTTGGTGGAAATGSGTGGGATCTCTGAAATGGAGACTCAATATGTATACCTTGACAMG	372
Db	479	TGGGTCGGAAACAATGACACAGAACTACAAACACGAAACCAAAATGTATATCTTGACACAA	538
Oy	373	TCAGATTC---TGGTGGACCGTATGTTTTCCTTCTTCCGATCGTTGAGAGGACCGTTCCGA	429
Db	539	AAGCACTCCGCGAGACCGACCTTATGTCTTACCTCTCCCAATCTCTGAAACACCTTCGGA	598
Oy	430	ACCTTCATTTCAAGCTGGGGATGATGACTTTTTCATGTTTGTCTCGAGAGTGTTCCGTCG	489
Db	599	ACCTCACTCCAAACCCGCTCACAACGATTCACATAGCAGTCTCCGTGGAAGGCGTTTCAACA	658
Oy	490	AAATCTTTGATGCATCGTTCCGAAAGTATGTTGTAATCTTCAATCGTGAGATGATCCGTTT	549
Db	659	CATGTCCACGGGTCCAAAGCTTCAAAAGCATGTCTTTACATCCATCTCAGTAAACAGCCCATAC	718
Oy	550	GCACCTTTAAAGAGCGGATGAAGATCGTAGAGACCCATCTTGAACCTTTCCGCTTGTG	609
Db	719	AGTATCTAAAGAAAGACAGATTAAAGTAATCCAACACTCATAGTAAGAACTGCTT	778
Oy	610	GAGGAGAAAGCTCCACAGGTATCGTGGACAAATTCGGTTGGTGCAGCTGGAGACCGCTT	669
Db	779	GAAACAAAACAGACCTATATATTATAGCAAAATTCGGTTGGTGCAGCTGGAGATGCTTTT	838
Oy	670	TACCTCAAGGTTTCATCCACAGGGCGTAAATAGAGCGGTGAGCATCTGTCGACGGCGGT	729
Db	839	TACTTGAAGGTTTCATCCAAAGGTGATGGAAGGTTGAAAGTCTCTCACAGATGTGGT	898
Oy	730	TGTGCTCCCGGTTTATGTCTTAATCAGCAGATGTTGGCAATTCATTCGGACACAGATTCGGAT	789
Db	899	TGTCTCTCCGGTTTCTGATTAATTCAGCAGCGGTGGCAATTCATTTGTCTATGACGATGAC	958
Oy	790	CCCATCAACCAAGAGGAATGAACCAACCGTGTGCGCGAGCAAAATCCCTGCCGCTT	849
Db	959	GATGAAATGATTTACGAAATGAACCGCAACTCCGCCGGGGAACAAATGCAATGCAAGACTT	1018
Oy	850	TTTGAATTTCCAAGAGAAATTCACAAATTCGCTGACTACGTCAATCCCAAGGCCACCGGCCCC	909
Db	1019	GTTAAATATCAGCAAGCAATTTCTAAGTTTAAAGAAATATGAAATCTTCGA-----A	1066
Oy	910	CGACCCGCGGAGAAAGGGGATGAAGCGGCTTTATAGATGAACCTCAAGAGAGGATTTAAAGCT	969
Db	1067	AATGAGAGGAAGAAAGSTTTGGGCTGTTTGTGTGAGGAGATTTAAGAAAGATTTGGGAGT	1126
Oy	970	GTTGAGAGCATGTTATGTTTGGCATGCTTTGTGTGATATTTGGGATGCGCTTCCGCCGAC	1029
Db	1127	GTTGAGAGTGTATATGTTTATGTTTGGCATGCGCTTTGTGGGATATTTGGGCGGGGTTAGCGTGA	1186
Oy	1030	GTTGCGCTGGCTTCCCTAGAGGACGTGTGATTCAGCCAGTCTTTCACACAGGCTGCAAGTG	1089
Db	1187	GTTGATGGGATGCCGAAACCTAGAGTTGTGTTGTCCGAAGTGTCTCAAGGGGTGAAGATG	1246
Oy	1090	ACGATGAGAGATTTGGCGGTGATGAATGTTTCTTCAATAAAGTGGCGCTGATCCCGCG	1149
Db	1247	ACGATGAGAGATTTGGGCGGTGATGAATGTTTGAAGACGCTGTGGGCTAGTGGCGCA	1306
Oy	1150	GAGAAAGCTGAGAGATGTACGAAGACATTCACTCATTTTGGAAAAAGTTGGGATCGAC	1209
Db	1307	GATTTTCCACATGAGATGTTGATGGGCTTCACTCTCATTTTGGAAATCGCGCGGAATATGAC	1366

QY	1210	GGTGTAAAGATTGACGCTTATTCACCTATTGGAGATGTTGTGGAAAGCATATGGAGGAGA	1263
Db	1367	GGGTGTTAAAGTGTGATTATTCACATCTGTGGAGTACTATCAGAGGAATATGTGTGACGA	1426
QY	1270	GTGGTATTGGCAAGGCAATTTACAAAGCAATGACCAAATTCATTAATTAACATTTTAAA	13239
Db	1427	GTTHAGCTAGCAGAGGCTTATTTACAAAGCACTAACCTCATAGTGAAGAAACATTTTCAA	1486
QY	1330	GGAATGGAGTCAATTGGCAACGTATGAAACATTTGTAACGACCTTATGTTCCCTGGACGAA	1389
Db	1487	GGCAATGGTGTAAATGCTAGACATGGAGACATTCGAACGACTCTTTCCTCCGACACGAA	1566
QY	1330	GCTATCTCTTGGTGTGTTGGTATGACTTTTGGTGCAGGAGACCCCTGTGTATCCA	1449
Db	1547	GGCATATCCCTCGGCGCCGTCGGAGATGATTTTGGTCTGTATCCATCTGTGTATCCA	1606
QY	1450	AAGGTATGTTTTGGCTCCAAAGGATGACATGAGTTTCAATTGGCCACAGACGCTTGGG	1509
Db	1607	AATGTCATATTTGGCTCCAAAGGTTGTCACATGATGTACATTTGGCTTACACACGTTTATGG	1666
QY	1510	ATGGGGAACCTTCATCCACCCTGCATGGGATATTTGTAATCCACACCACCCTGTGTCGCC	1569
Db	1667	ATGGGAATTTTCATTCAGCCACAGATTTGGGACATGTTTCACTCCACATCATCTTGTCTGAA	1728
QY	1570	TTTCATGCTGCTCTCGAGGCCATCTGTGTGGCCGATCTATGTAATGATTTCTGTGGA	16239
Db	1727	TTTCATGCGCGCTCACAGGCCATATCCGGGCGACCAATTTATGTAATGATTTGTGTGT	1786
QY	1630	AAGCATTACTTTGATCTTCTGGAAAAAACTAGTCTTCTGATGATCGATCCTTGTGAAGT	1688
Db	1787	AATACATAATTTCAAGTGGCTCAATATCTTTGTTTTCCCGATGCTTCATATCTGTGGTGT	1846
QY	1690	GAGTACTATGCACTCCCGACCTCGGATGTTGTTTGAAGACCCCTTGGCATTAATGGAGA	1749
Db	1847	CAACATTAGCGACTCCCTACAAGAGATGCTTGTTTGAAGACCCCTTGGCATTAATGGCAA	1906
QY	1750	ACTATGCTTAAAGTTTGGAACTCTCAACAGTTTCACTGAGATGATTTGTGCATTCAATGTC	1809
Db	1907	ACATATGCTGAATAATTTGGATCTCAACAATAATACAGGTGTTTGGCTTTTCAACTGC	1966
QY	1810	CAAGAGAGAGATGTGCTCGTAGACACCGCCGACCAATGCTTTTCACAATATCTCAAA	1866
Db	1967	CAAGTGTGTGGTGCTCTCGTAGGACACGGCGAACAAGTATCTGAAATTTTTCACGC	2026
QY	1870	CGAGTACATCAACAACTTAACCCAAAAGCATATGAATGACAGTGGAGAAACCCATTC	19239
Db	2027	CGGTCGACATCTTATGCAAGTCCCGAAGACATTTGAATGTCGCAATGGGAAACCTCAATG	2086
QY	1930	TCATATGAAGCGTTAAACCTTTGGCGTTTACCTCTATCAAGCCAAAACCTTATCTC	1988
Db	2087	AGCACCAAGGTGTGGATTTTTTTTGGTGTTATTTTTCAAGAGAGAAATTTGAGGCTC	2146
QY	1990	TCCAAAGCCCTCAAGATCTTTGACATAGCTCTTTGACCCTATGGAATTCAGACTCATCT	2043
Db	2147	ATGAAAGTGTCTGTATAGATGAAAGTTTCCCTTGACCCATTTAGTTTGAAGCTAATGACA	2206
QY	2050	GTTTTACCAGTGAACAAATCATCAACATCTCTACATCTTTGGCCCATTTGGGCTGGG	2109
Db	2207	GTGTCTCCAGTGAAGTGTTTTGGAAAAGTTTATACAGTTTTCACACCAGTATGGCTTACTG	2266
QY	2110	AACATGCTTAACTACTAGTGAAGCCATCCATCTGTGGACTATGACAGATGACCTTAAGTCA	2163
Db	2267	AACATGCTGAACCTGTGTGGGTGGGATTCAGTCTCTGGAGCTTTGATGATATATGCAAGTTGG	2326
QY	2170	GTCGAGATTTGCTCAAAAGGCTGTGTGATAGTCGAGTATTTGTCATCGAAAACCAAGG	2223
Db	2327	GTCAGAAATTTGGGTGGAAGAGGTTCGGGGAGATGAGCGTGTTCGCTGAGAAACCGGTT	2386
QY	2230	GCTTGTGCTATTAATGATGGGAGAGATCTTTGGGCTCAACTATGATCAGAGAACCAATGTGTGT	2289
Db	2387	TGCTGCAAAATTTGATGGGTTTTAAAGTGAAATTTCTTTATG---AGACAAAAATGGCAGAA	2443
QY	2230	GTTCAGAGTGCATGGCCAAATTTGATCTTCTCATCG	2322

Qy	1513	GGGAACTTAAATCCACCCTACAGGGATATGTTCCAAATCCAAACCAACCCCTGGGCGCCTTC	1512
Qy	1513	GGGAACTTAAATCCACCCTACAGGGATATGTTCCAAATCCAAACCAACCCCTGGGCGCCTTC	1512
Db	1468	GGCAATTTATTCACCCAGATTTGGGACAATGTTCCAAATCTACTCAATCCCTTGGTGCCTTC	1527
Qy	1573	CATGCTGCTCTGTGAGCCATCTCTGGTGGCGGCAATCTATGTTAGTGAATTCGTGTGGAAAG	1632
Db	1528	CATGCTGCTCTGTGAGCCATCTCTGGTGGCGGCAATCTATGTTAGTGAATTCGTGTGGAAAG	1587
Qy	1633	CATTAACCTTTGATCTTTCTGAAAAAACTAGTGTCTTCTGATGATGATGATCTTTCGAATGAG	1652
Db	1588	CACAACTTTTACGTGCTTTAAGACCTTTGGCCCTTGGCCAGATGGGTCCATCTCAGATGTGAG	1647
Qy	1693	TACATCTACATCCCGACTGGCGATTTGTTTGTGAAAGACCCCTTTCATTAATGGAANAAT	1752
Db	1648	CACATCTACATCCCGACTGGCGATTTGTTTGTGAAAGACCCCTTTCATTAATGGAANAAC	1707
Qy	1753	ATGCTTAAGATTTTGGAAATCTCAACAAGTTTCACTGAGTAGTATGTGATTCACATGCGCAA	1812
Db	1708	ATGCTCAAGATATGGAACATCAACAAGTACACTGGAGTTCTTGGGGTGTTTAACTGCGAG	1767
Qy	1813	GGAGAGAGATGCTGTGCTGTGAGACAGCCCGCCCAACCAATGCTTTTTCACATTACTCAAAAG	1872
Db	1768	GGAGAGAGTTGGTTCGTCGTGAGATTAGTGTCCAAACAATGTGCTGCTGAATTTTTCATAGG	1827
Qy	1873	GTCGACATCCAAAACCTAAACCAAGAATAGATAGATGGACAGTGGAGAAAACCCATATCTCT	1932
Db	1828	GATATCAACCAAGACCAATATACAAAGACATTTGAATGGGATAGTGGAAAGAAATCCATTTCC	1887
Qy	1933	ATTGAAAGGCTTAAACCTTTGGCGCTTTACCTCTATCAAGCCCAAAAACCTATCCTCTCC	1992
Db	1888	ATTGAAAGGCTTGAACCTTTGCGCTGTGATTTTCAGGCAAGCCCAAAAACCTATCCTCTCA	1947
Qy	1993	AAGCCCTCTCAAGA - - - TCTTGAATAGCTCTTGACCCATTCGATAGTTGAGCTCATCTCT	2049
Db	1948	GCACCACTCTATGCACTGAGAGATTTCTCTTGGAGCCCAATTAATTTGAGCTTATTAACA	2007
Qy	2050	GTTTTCACACAGTACCAAACTCAATCCAAACCTTCTCTACATTTGGCCCAATGGCGCTGGTG	2109
Db	2008	GTTTTCACACAGTACCTGCTTCCCTGGCAGTACGTAGTAATTTGCTCTATTTGGTTGGTG	2067
Qy	2110	AACATGCTTTAACTAGTGTGAGCCATTCATCTGTGTGACTATGACGATGACCTTAAGCTCA	2159
Db	2068	AATATGCTTAAACACTGTGTGTGAGCAGTCCATGCTTTTGTGATGAGAGGCTCAGAAATTTG	2127
Qy	2170	GTCAGAGATTGGTGTCAAAAGGCTGTGTGTGAGATGCGAGTATTTTGCATGCAATGCAAAAG	2229
Db	2128	GTTTAAAGTTGGTTTAAAGGCACTGGGAGATGAGAGTCTTATGCTTCAAGACCAAGCA	2187
Qy	2230	GCTTGTGCTATTTGATGGGAGAGATGTTGGGTTCACAGTATGATCAGACCAATGCTGTGTG	2289
Db	2188	ACGCTGTGAATTTGATGTGCAAAAGATTTGATTTTGAATATG - - - AAGGGCTATGTGTCAAC	2244
Qy	2290	GTTCAAGTGCATGCGCAATTAATGATCTTCTCATCG	2322
Db	2245	ATTCAAGTACCAATGCGCTGTGCTTCTCAAAATTTG	2277

RESULT 12	
AAV40802	
ID	AAV40802 standard; cDNA to mRNA; 1762 BP.
XX	
XX	
AC	AAV40802;
XX	
DT	23-SEP-1998 (first entry)
XX	
XX	
De	Japanese artichoke raffinose synthetase coding sequence.
XX	
KW	Raffinose synthetase; metabolism modification; food additive;
XX	gastrointestinal flora; Japanese artichoke; ss.
XX	
OS	Stachys sieboldii.
XX	

Key	Location/Qualifiers
FT CDS	2..1762
FT	/*tag= a
FT	/product= raffinose synthetase
FT	/note= "no stop codon given"
XX	EP849359-A2.
XX	
PD	24-JUN-1998.
XX	
XX	18-DEC-1997; 97EP-0122417.
XX	
PR	18-DEC-1996; 96JP-0338673.
XX	
PA	(SUMO) SUMITOMO CHEM CO LTD.
XX	
P1	Oeda K, Watanabe E;
DR	WPI: 1998-324670/29.
XX	
DR	P-PSDB: AAM57888.
XX	
PT	New nucleic acid molecule encoding plant raffinose synthetase -
PT	capable of producing raffinose, used as food additives with
PT	beneficial effects on gastrointestinal flora
XX	
PS	Claim 1; Page 36-38; 44pp; English.
XX	
CC	This sequence encodes the Japanese artichoke raffinose synthetase of the
CC	invention. The raffinose synthetase is capable of producing raffinose by
CC	combining a D-galactosyl group through an alpha (1-6) bond with a
CC	hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC	residue in a sucrose molecule. The DNA can be used to modify metabolism
CC	of a host organism by introducing into the host organism or its cell so
CC	that the content of the raffinose family oligosaccharides in the host
CC	organism or cell is changed. Raffinose oligosaccharides are useful as
CC	food additives with beneficial effects on the gastrointestinal flora.
XX	
SO	Sequence 1762 BP; 395 A; 451 C; 499 G; 417 T; 0 other;
XX	
Query Match	37.7%; Score 886.8; DB 19; Length 1762;
Best Local Similarity	70.0%; Pred. No. 8.7e-275;
Matches 1223; Conservative	0; Mismatches 517; Indels 6; Gaps 2
OY	324 A A A T G T G G G G A T C T T G A A T T G G A G A C T C A G A T T G T G A T C T T G A G A A G T C A G A T T T G G 383
DB	4 A A A C G G G T C G C A T C T T G A C G G G A A A C T C A A T A T A G C G T C G A C A A G T C C G A ---CGA 60
OY	384 T G C A G C G T A T G T T T T C C T T C C C A T C G T A A G G A C C G T T C C G A A C C T G A T T G A G C C 443
DB	61 C A G G C C T T A C A T C G T G C T G C T T C C G T C A T C A G G G G A G A T T T G G G C T T C C C T T A G C C 120
OY	444 T G G G A T G A T G A C T T T G T G C A T T T G T G C A G A T G T G T T G C T G A A A G T T G T T A T G C 503
DB	121 C G G T G C A T G A T T T A T T A T T G T G T C G A A A C G G G T C A A C C A A G T C A A G A G T C 180
OY	504 A T C G T T C C A A G T A T G T T A T C T T A T C T G T G A T T C C G T T T G C A C T T G T T A A G A 563
DB	181 C T C G T T C C T G C T T G C T C T A C A T G C A C C C G G T G A T G A C C T T T A C C T G G T G A A G A 240
OY	564 G G C G A T A G A A T C G T G A G A C C C A T T T G A A C T T T T C G T T T T G T T G A G A A A G A C T C 623
DB	241 C G C C G T G A A G T G G C C G C C A C C A C C T C G G A C G T T C A G G C T C T G A G A A A A A C T C C 300
OY	624 A C C A G A T A C T G G A C A A A T T C G G T G T G T C A C G T G G A C G C G T T T A C T T A A C G G T T C A 663
DB	301 G C C G G G A T C G C A C A A A T T C G G G T G T G A C A C G T G G A T T G C G T T A C C T C A A C G T C C A 360
OY	684 T T C A C A G G G C G A A T T G A A G C G T G A G C A T C T C G T C G A C G G G G T T G T C T C C C G G T T T 743
DB	361 G C C C C A C G G C G T T A T G A A G G G C T G A G G G G C T G T T G A C G G G G A T G T C C G C G G G C C T 420
OY	744 A C T C C T A A T T C G A A G A T G T T G C A A T C C A T T C G A C A C A G A T T C G A T C C A T C A A C A A G A 803


```

Db 421 GGTGTTGATCGACGAGGGTGGCATTCATTGTCACGACACGAGCGCTCACACCGA 1480
OY 804 AGGAATGAACCAACCCGTGCGCGAGCAAAATGCCCCGTCTTTTGAATTCACA 863
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 GGGATGGGAGAACCTCCGCCGAGAGCAAAATGCCCTGAGGTATCATCAAGTTGAGGA 540
OY 864 GAATTCAAATTCGCTGACTGATCAATCCCAAGGCCACCGCCCCGAGCGGCCAGAA 923
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GAATTCAGATTCAGGAGATGACGAGAGCCGAATAAAAGTGGGCC--GGGCCCGAATAC 597
OY 924 GGGAGTGAAGCGCTTTATAGATGACATCAAGAGAGTTTAAGATGTGGAGCATGTTTA 993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 GGGGATGGGGGCTTTATTCGTGATCAATGAAGCAATTTCAAGATGTGACTACGTGA 657
OY 984 TGTTCGATGCTTTGTGTGATATTTGGGTGGCTTCGCCCGCAGGTGCTGCTTGGCC 1043
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 658 CGTGTGATGCTGTTGTGTGATTTGGGGCGGGCTCAGGCCCAATGTTCCGGGCGCC 717
OY 1044 TTAGGCACTGTGATTCAGCCAGTGCCTTCACAGGGCTGACATGACATGAGGATTT 1103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 CGAGGCTAAGCTCATTCAGACCCAACTGACTCCTGGGCTTAACACCACTGAGAACTTT 777
OY 1104 GCGGTGATGAATGTTGTTCTTCAAGTGGGCTGGTCCCGCGAGAGGCTGAGGA 1163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 GGTGTGTGATGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
OY 1164 GATGTACGAGAGCACTTCATGCTCATTTGGAAAAAGTTGGGATGACGCTTAAAGATTGA 1223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 AATGTATGAAGGATTCATTCATCATCTCGAATCTGTGGGATGTGATGATGATGATGATGAT 897
OY 1224 CGTATCCACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 898 CGCATTCACATTTGTGTGAATGTGTGTGAAGACTATGTGTGTGTGTGTGTGTGTGTGTGT 957
OY 1284 GGCATTCACAAAGCAATGACCAATCAATTAATAAATTTTAAAGAAATGAGTCAAT 1343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 958 GGTATTTCAAGGCTTATCAAGCTCAGTATACAACTTCAAGGCAAGCGCTCAT 1017
OY 1344 TGCAGTATGAGCAATGTTGACGACTTCAATGTTCTTGGCAGGAAGCTATCTCTTGG 1403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1018 CCGTGCCCTGGGACATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1077
OY 1404 TCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1078 TCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1137
OY 1464 GCTCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1138 GTTGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1197
OY 1524 CCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1198 CCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1257
OY 1584 TCGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1258 ACAGGCAATGCTCGGGGGGCCCATTTACATGATGATGATGATGATGATGATGATGATGATGAT 1317
OY 1644 TCTTCGAAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1318 GCTCCTTAGAGCTGCTCTTCCGATGCTCCTCCGTTGATGATGATGATGATGATGATGATGAT 1377
OY 1704 CCCGACTGCGATGTTGTTTGAAGACCCCTTGTGATTAATGAGAAATGATGATGATGATGAT 1763
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1378 TCCGACTCGATGCTCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1437
OY 1764 TTGGAATTCACAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1823
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1438 TTGGAATTCACAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1497
OY 1824 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1883
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1498 GAGCGGGGAGATGCTGCGAACCAATGCGCTGCGGAGTATTCGCCAGCGCTCTCTTAG 1557

```

```

OY 1884 AACTAACCCAAAAGACATAGATGACACAGTGGAGAAAACCTTATCTTATGAGGCGT 1943
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1558 CGGTGGTCCAGTACATTCATGATGAGAAAGCAAGACAGATCCGATGACGTGACGGCGT 1617
OY 1944 TAAACCTTTGGCGCTTTACTTATCATAGCCAAAACCTTATCTCTCCAGCCCTCTCA 2003
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1618 CAACACATTCGCGTGTACTTATTCACGAGAAAGAACTGCTCTTCTTAAGCATTCAGA 1677
OY 2004 AGATCTTGACATGATGCTCTTACCATTCGAAATTCGAGCTTCATCAGTGTTCACACAGTAC 2063
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1678 CAACATTCGACATCAGCTTACGCTTACGCTTTCGATTTGAGCTATACCGTTTCTCAGTCAA 1737
OY 2064 CAACCT 2069
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1738 AACTCT 1743

RESULT 13
AA210003
ID AA210003 standard; cDNA to mRNA: 1762 BP.
XX
AC AA210003;
XX
DT 26-OCT-1999 (first entry)
XX
DE Nucleic acid encoding a raffinose synthase protein.
XX
KW Raffinose synthase; plant; sucrose; raffinose; ss.
XX
OS Stachys sieboldii.
XX
FH Key 2.1762 Location/Qualifiers
FT CDS /tag= a
FT FT /product= "raffinose synthase"
FT FT /note= "no termination codon given"
XX
PN JPI1215984-A.
XX
PD 10-AUG-1999.
XX
PF 12-DEC-1997; 97JP-0342899.
XX
PR 28-NOV-1997; 97JP-0329006.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SDMO ) SUMITOMO CHEM CO LTD.
XX
DR WPI: 1999-511112/43.
DR P-PSDB: AAY30144.
XX
PT New raffinose synthase gene - is prepared from a plant material
XX
PS Claim 13: Page 31-34: 40pp: Japanese.
XX
CC The present sequence encodes a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6)-D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
SQ Sequence 1762 BP; 395 A; 451 C; 499 G; 417 T; 0 other:
XX
Query Match 37.7%; Score 886.8; DB 20; Length 1762;
Best Local Similarity 70.0%; Pred. No. 8.7e-275;
Matches 1223; Conservative 0; Mismatches 517; Indels 6; Gaps 2;
OY 324 AAATGGTGGGATCTTGAATCGAGACTCAGATTGTGATCTTGAGACTCAGATTCTGG 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 AAACGGGTGCGATCTTGACCGGGAACCTCAATATAGCTGCTGACAGATCCGA--CGA 60
OY 384 TCGACGTAATGTTTCTCTCTTCCGATTCGTGAGGAGCCGTTCCGAACCTCGATTGAGCC 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 61 CAGGCGCTACATGCTGCTTCCGCTCATCGAGGGGAGATTGGGGCTTCCCTTCAGCC 120
Oy 444 TGGGATGATGACTTGTGATGTTGTGTCGAGAGTGGTGTGTCGAAAGTTGTGATGC 503
Db 121 CGGTGGGATGATTTATGATATTTGTGTGCGAAGCGGGTCAACCAAGTCAACAGATC 180
Oy 504 ATGCTTCCGAAAGTATGTTATCTTCATGCTGGTGATGATCCGTTTGCATTGTTAAAG 563
Db 181 CTGCTTCCGCTGCTGCTCTACATGCAACGCCGTGATGACCTTTTACCTGGTGAAGGA 240
Oy 564 GCGGATGAAGATGCTGAGAGACCATTTGGAACTTTCCGTTGTGGAGAGAAGACTGC 623
Db 241 CCGCGGAGAGGGGCGCGCCACCACTTCGGGAGCTTCAGGCTGCTGGAGAGAACTCC 300
Oy 624 ACCAGATATCGTGACAAATTCGGTTGGTGACGTGGGAGCGGCTTTACTTAAAGGTTCA 683
Db 301 GCGGGGATTCGTCGACAAATTCGGGTGGTGACGTGGATGCGCTTTACTTCACTCACTCA 360
Oy 684 TCCACAGGGCGTAATGAGAGCGGTGAGGATCTCTGACGCGCGGTTGTCTCCGTTT 743
Db 361 GCGCCAGCGCGTATGAGAGGGGTGCGAGGGGCGTGTGACGGGAGATGTCGCGCGGGGCT 420
Oy 744 AGTCTTAATCGAGATGTTGGCATTCATCGGACACGATTCGATCCATCAACAAAG 803
Db 421 GGTGTGATCGACGAGGGGTGCGAGTCTTGTACGACAAACGACGCGCTCACACCGA 480
Oy 804 AGGAATGAACCAACCGGTGCGCGGAGCAAAATGCCCTGCCCTTTTGAATTCACAAG 863
Db 481 GGGGATGGGAGACACTCCGCGGAGAGCAAAATGCCCTGACGTTATCAAGTTTGAAGA 540
Oy 864 GAATTAACAATTCCTGACATCTCAATCCCAAGGCCACCGGCCCGGAGCCGCGGAGAA 923
Db 541 GAATTACAATTTCAGGAGACGAGAGAGCCGAATTAACCTGGGCGC---GGGCGCGAATAC 597
Oy 924 GGGGATGAAGCGCTTATATGATGACTCAAGAGAGATTGAAGCTGTGAGAGATTTA 983
Db 598 GGGGATGGGGCGCTTATCTGTCATGACATGAAGCAATTTCAAGAGTGTGACATCACTGA 657
Oy 984 TGTGGCATGCTTGTGTGATATTTGGGGTGGCTTCCCGCCGACGTCCTGGCTTGGC 1043
Db 658 CGTGTGCAATGCGCTTGTGTGATTTATGGGGCGGGCTCAGGCCCAATGTTCCGGGCTGCC 717
Oy 1044 TGAAGCAGCTGTGATTCACCCAGTCTTTCACCCAGGGCTGACAGATGAGTGAAGATTT 1103
Db 718 CGAGGCTTAAGCTATGAGCCCAACTGACTCCTGGGCTTAAAGACCACTATGAGAAATTT 777
Oy 1104 GCGGATGATGAATGTGTCTTCTCATAAAGTCCGGGCTGTGCTCCCGCCGAGAGCTGAGA 1163
Db 778 GCGTGTGATGAAGATTGTCAACAATGGCGTGGGTGTGCTCCACCGAGTGTGTTGAACA 837
Oy 1164 GATGTGAGAGAGACTTCATGCTCATTTTGGAAAAAGTTGGGATGAGCGTGTGAAGATTGA 1223
Db 838 AATGTATGAAGGATTTACATTCATCTCGAATCTGTGGGATTTGATGAGTCAAAAGTTGA 897
Oy 1224 CGTTATCAACCTATTTGAGAGATTTGTGTAAGACTATGAGAGAGAGATTTGGCAAA 1283
Db 898 CGTCACTCAATTTGTGGAATGTTGTGAAGACTATGAGGAGAGATGACTTACCCAA 957
Oy 1284 GCGATTTTCAAAAGCAATGCAACCAATCAATTAATTAACATTTTAAAGGAATGAGTCAAT 1343
Db 958 GCGTATTTCAAGAGCGCTTATCAAGCTCAGTTAAACAACCACTTCAAGCGCAACGCGTCAAT 1017
Oy 1344 TGCAGATATGAGCAATTTGACGACTTCAATGTTCTTGGCAGGAGATATCTCTGTGG 1403
Db 1018 CCGTGGCTTGGAGCACTGCAATGACTTCAATGTTTCTCGAAGCGAGGCCATTAACCTTGGG 1077
Oy 1404 TCGTGTGATGATGATTTTGTGACGAGACCCCTCTGTGATCAACAAGCGATTTTGG 1463
Db 1078 TCGTGTGCGGAGATGATTTTGTGCTGACTATCTGAGATCCCAATGGCAGCTTCG 1137
Oy 1464 GCTCCAAAGATGTCANATGTTTCAATGTTGCGCAAGCAGCACTTGTGATGGGGAATTCAT 1523
Db 1138 GTTTCAAAGGCTGTCATGCTGCACTGCGCTTAAACAACAGCATATGATGGTAAATTTCAAT 1197

Oy 1524 CCACCTGACTGGGATATGTTCCATTCACACCCACTTGGCGGCTTTCATGCTGCTC 1583
Db 1198 CCACCTGATTTGGAGATATGTTTCAATTCAGACTCAACCTTGGCGGATTTCCAGCTGCTC 1257
Oy 1584 TCGAGCATCTGTGTGCGCCGATCTAGTTAGTGAATTTGTGGAAAGCAATATTGTA 1643
Db 1258 ACGAGCATCTCGCGGGGCGCCATTTAGCTAGCTGACTGCTGGTGGAAAGCACAACTTGA 1317
Oy 1644 TCTTTCGAAAAAACTAGTCTTCTCATGATGATTCCTTCGAAGTGAAGTACTATGACT 1703
Db 1318 GCTCTTAGAGCGCTGTTCTTCCGATGCTTCATCTCCGTTGTGATTAATAAGCGCT 1377
Oy 1704 CCGACTCGCATGTTGTTTGTGAAGACCCCTTTCATATGAGAAAGTATGCTTAAAGAT 1763
Db 1378 TCGGACTCGCATGCTGCTCTTTGAGATTCACATTCACAAATGCAAGACTATGCTCAAAAT 1437
Oy 1764 TTGGAATCTCAACAGTTCACTGAGTGAATGGTGCAATTCACATGCGCAAGAGAGAGATG 1823
Db 1438 TTGGAATTTAACAAGTTCAACCGAGTTGTGGAATCTTCAACTGCGCAAGGTGGCGGTG 1497
Oy 1824 GTGTCTGAGACACCGCGCAACCAATGCTTTTACAAATACTCAAAAGAGTGCATCA 1883
Db 1498 GAGCCCGGAAGTCCGTGCGAACCAGTCCGCTGCGAGTATTCACAGCGCGCTCTAG 1557
Oy 1884 AACTAACCCAAAGACATGAAATGAGCAGAGTGAAGAAACCCATCTCTATTGAAGGCGT 1943
Db 1558 CGTGTCTCGAGTGAATGATGAGTGAAGCAAGAGACAGATGCGATGACGTGACGCGCT 1617
Oy 1944 TAAACCTTTGCGCTTACCTCTATCAAGCCAAAAAATTAATCTCTCCAGAGCCCTCTCA 2003
Db 1618 CAACATTCGCGCTTACTTATTCACGAGAGAAAGTCTGCTTTTCTAAGCATCAAGA 1677
Oy 2004 AGATCTTGAATGCTCTGACCATTCGAAATTCGAGCTCATGCTTTCACCAATGAC 2063
Db 1678 CAATAATGCAATCACGCTTGAGCCCTTTCGATTTGAGCTGATTAACGCTTTCACATCAA 1737
Oy 2064 CAACCT 2069
Db 1738 AACTCT 1743

RESULT 14
AA220210
ID AA220210 standard: cDNA: 1762 BP.
XX
AC AA220210;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rapeseed raffinose synthase cDNA.
XX
KW Raffinose synthase: rapeseed; transgenic plant; ss.
XX
OS Brassica napus.
XX
FT Key
FT CDS 1.1719
FT Location/Qualifiers
FT /tag= a
FT /note= "this region is specifically claimed in
Claim 9"

EP953643-A2.
03-NOV-1999.
27-APR-1999: 99BP-0107430.
30-APR-1998: 98BP-0120550.
30-APR-1998: 98BP-0120551.
04-DEC-1998: 98BP-0345590.
10-DEC-1998: 98BP-0351246.

Db 1603 GTGAGCTGCAGATGATGATGAAGATGTTGAGTTGGTAGC---AAGAGTCAGTG 1659
QY 2287 GTGCTTCAAGTCCATGCC 2306
Db 1660 ATGGTTCAAGTGCCTTGGTC 1679

RESULT 15
AA220207
ID AA220207 standard; cDNA; 928 BP.
AC AA220207;
XX
DT 17-JAN-2000 (first entry)
XX
DE Soybean raffinose synthase cDNA.
XX
KM Raffinose synthase; soybean; transgenic plant; ss.
XX
OS Glycine max.
XX
FH Key Location/Qualifiers
FT CDS 2..799
FT /tag= a
XX
PN EP953643-A2.
XX
PD 03-NOV-1999.
XX
PE 27-APR-1999; 99EP-0107430.
XX
PR 30-APR-1998; 98JP-0120550.
PR 30-APR-1998; 98JP-0120551.
PR 04-DEC-1998; 98JP-0345590.
PR 10-DEC-1998; 98JP-0351246.
XX
XX (SUMO) SUMITOMO CHEM CO LTD.
XX
XX Watanabe E, Oeda K;
XX
XX MPI: 1999-593144/51.
XX P-PSDB; AAY32072.
XX
XX New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
XX
PS Claim 3; Page 20-22; 55pp; English.
XX
XX This is the nucleotide sequence of a soybean cDNA clone coding
CC for raffinose synthase (see AAY32072), a protein which can bind a
CC D-galactosyl group through an alpha(1-6) bond to the hydroxy group
CC attached to the carbon atom at the 6-position of the D-glucose
CC residue in a sucrose molecule to form raffinose. The cDNA was
CC isolated from soybean cv. Williams 82 immature seed cDNA by PCR.
CC Probes or primers generated from plant raffinose synthase genes
CC (see AA220207-10) may be used to obtain other raffinose synthase genes
CC by labeled detection or amplification (claimed). These genes may
CC be used to control the levels of raffinose produced in plants.
CC Antisense genes can be used to knock out existing gene activity,
CC and sense genes to increase the level of gene activity. The
CC resulting transgenic plants may be used as a food source to alter
CC the growing conditions for gut enterobacteria, providing general
CC health advantages. The sense and antisense genes may also be used
CC in gene and phenotypic analysis of plants and for the selection of
CC plant clones with the desired characteristics with respect to
CC raffinose content.
XX
SQ Sequence 928 BP; 266 A; 171 C; 222 G; 269 T; 0 other;

Query Match 12.9%; Score 303.8; DB 20; Length 928;
Best Local Similarity 63.0%; Pred. No. 8; 9e-87;
Matches 486; Conservative 0; Mismatches 282; Indels 3; Gaps 1;

QY 1545 CCAATCCACCACCCCTTGTGCGCCTTCCATGCTGCGCTCGAGCCATGCTGTGGGCC 1604
Db 1 CCAATCTGATCATGCTTGTGCGCAATTTCCAGCGTCTTACAGCCATTTCTGTGGACC 60
QY 1605 GATCTATGTTAGTATGATTCGTGGAAAGCATTAATTGATCTTCTGAAAAAATAGTGT 1664
Db 61 AATTTATGTAAAGGACCTGTGTTGAAACACAACTTCAAGTTGCTTAAGAAGCTTGTTC 120
QY 1665 TCCTGATGATGATCCTTTCGAAGTGAATGATGATGATGATGATGATGATGATGAT 1724
Db 121 ACCGTATGCTGCTCATTGTTGCGGTGTCACATATGATGATGATGATGATGATGAT 180
QY 1725 TGAAGACCCCTTGTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1784
Db 181 TGTAGATCTCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 1785 TGGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1844
Db 241 CGGGTTTGGGTGCTGTGTTCAATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 1845 CCAATGCTTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1904
Db 301 CAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 1905 ATGGCAGTGGAGAAACCCCTATCTCTATGTAAGGCGTTAAACCTTGGCGTTACCT 1964
Db 361 ATGGGCAAGAGGAGACACCCAGTTGATGATGATGATGATGATGATGATGATGATGAT 420
QY 1965 GTATCAAGCCAAATTAATCTCTCAAGCCCTCTCAAGATCTTGATGATGATGATGAT 2024
Db 421 GTTAAGAGACGACAAAGTTGAAGCTGCTGAAGATGATGATGATGATGATGATGATGAT 480
QY 2025 CCAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2084
Db 481 GCGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 2085 ACACTTGGCCCCAATTTGGGCTGTGAACATGCTTAACACTAGTGAAGCCATCAATCTGT 2144
Db 541 CCAATTTGGCCCCAATTTGGGCTGTGAACATGCTTAACACTAGTGAAGCCATCAATCTGT 600
QY 2145 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2204
Db 601 GGAATTTGATCAACAGGAAATTTGGGCTGTGAAGATGATGATGATGATGATGATGAT 660
QY 2205 AGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2264
Db 661 GGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 2265 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2315
Db 721 TTATGTT---GATAGAACGTGAGGCTCCAGTCTGTGTGCTTCTTC 768

Search completed: April 3, 2003, 16:22:10
Job time : 627 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 16:04:40 ; Search time 3761 Seconds
(without alignments)
10128.100 Million cell updates/sec

Title: US-09-675-208-4_COPY_56_2407
Perfect score: 2352
Sequence: 1 ATGGCTCCTGAGTTTAA...CGTTATCGACTGTTT 2352

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlinu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	416	17.7	668	12	BF635461 NF080A03D
2	402	17.1	646	9	A1727515
3	371.6	15.8	649	12	BF636478 NF089A04D
4	363.2	15.4	616	10	BE131081 L48-1031T
5	339.6	14.4	515	9	A1727581 BNLCG1847
6	335.4	14.3	682	12	BF634464 NF061C04D

7	322	13.7	559	13	BI974727	BI974727 sa172D10.
8	304.2	12.9	564	12	BE269676	BE269676 GA_EB000
9	301.4	12.8	560	14	BM887415	BM887415 sam38B08.
10	295.6	12.6	489	10	AV421962	AV421962 AV421962
11	293.4	12.5	513	12	BF636592	BF636592 NF090H02D
12	274	11.6	818	12	BG321136	BG321136 Zm04_05d1
13	268.2	11.4	528	14	BM886707	BM886707 sam29B05.
14	268	11.4	525	12	BF425566	BF425566 su44c11.Y
15	266.6	11.3	530	9	AT993310	AT993310 701496022
16	264.4	11.2	435	12	BF633327	BF633327 NF054D03D
17	264	11.2	486	10	AW944715	AW944715 00168_1ea
18	259.2	11.0	714	14	BQ969695	BQ969695 QHB39E15.
19	251.2	10.7	591	12	BG043590	BG043590 su40f05.Y
20	248.8	10.6	695	12	BF633311	BF633311 NF059608D
21	237.8	10.1	672	12	BG454761	BG454761 NF104H06L
22	236.8	10.1	558	12	BG600526	BG600526 EST505421
23	236.6	10.1	435	10	BE248707	BE248707 NF010D10D
24	233.4	9.9	801	17	BH483423	BH483423 BOCG2056TR
25	211.2	9.0	518	10	AM649251	AM649251 EST327705
26	210	8.9	597	12	BG450458	BG450458 NF014C06D
27	203.2	8.6	832	14	BQ508902	BQ508902 EST616317
28	195.8	8.3	541	14	BM887020	BM887020 sam33C03.
29	190.8	8.1	714	14	C72550	C72550 C72550 RICE
30	188.6	8.0	592	12	BF632552	BF632552 NF027E01D
31	187.4	8.0	540	12	BG455277	BG455277 NF046E04P
32	185.6	7.9	375	10	BE187139	BE187139 NXNV_159
33	183.6	7.8	556	12	BF632381	BF632381 NF027B09D
34	181.6	7.7	557	12	BF480406	BF480406 LO-2338T3
35	181	7.7	496	9	AU083889	AU083889 AU083889
36	177.6	7.6	635	10	AV538951	AV538951 AV538951
37	174.2	7.4	411	14	T75887	T75887 10665_Lambd
38	170.4	7.2	610	12	BG648423	BG648423 EST510042
39	168.4	7.2	322	14	DA0721	DA0721 RICS2852A.R
40	166.4	7.1	671	12	BG452557	BG452557 NF107B10L
41	166	7.1	307	12	BF516701	BF516701 NX51_002
42	163.8	7.0	360	9	A1442026	A1442026 sa66f09.Y
43	162	6.9	446	10	BE346804	BE346804 sp31d06.Y
44	160.2	6.8	700	13	BI406578	BI406578 171E01 Ma
45	153.2	6.5	830	17	BH483417	BH483417 BOC2056TR

ALIGNMENTS

RESULT 1
LOCUS BF635461 668 bp mRNA linear EST 19-DEC-2000
DEFINITION NF080A03D1F1020 Drought Medicago truncatula cDNA clone NF080A03D1
5', mRNA sequence.
ACCESSION BF635461
VERSION BF635461.1 GI:11899619
KEYWORDS EST.
ORGANISM Medicago truncatula
SOURCE barrel medic.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 668)
Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
Unpublished (2000)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert length: 668 Std Error: 0.00
Plate: 080 row: A column: 03

FEATURES	Seq primer: TCACACAGGAACAGACTATGAC.
source	Location/Qualifiers 1. 668 /organism="Medicago truncatula" /db_xref="taxon:3880" /clone="NF080A03Df" /clone_id="Drought" /tissue_type="Plantlets" /dev_stage="Pooled timepoints" /note="Vector: Lambda Zap; Contains a mixture of entire platelets harvested in a series of days-post-watering timepoints."
BASE COUNT	181 a 125 c 148 g 214 t
ORIGIN	
Query Match	17.7%
Best Local Similarity	77.9%
Matches 514:	Conservative 0; No.M.1e-11; Indels 1; Gaps 1

	Query Match	Similarity	17.7%	Score 416:	DB 12:	Length 668:
	Best Local Similarity	77.9%	Pred. No. 1e-111:			
	Matches 514:	Conservative	0:	Mismatches 145:	Indels	Gaps 1:
OY	1155	GGCTGAGAGATGTACGAAGAAGACTTCA-TGCATTATTTGGAAAAAGTTGGATCGACGGTG	1213			
Dd	8	GGTTGATCAAAATGTATGAAAGGTATTCACCTCTCTTTGGAAAGTCTGGAAITGATGGGG	67			
OY	1214	TTAAGATTGACGTTATCCACCCTATTGGAGATTTGTGTGAAGACTATGGAGGAGAGTGG	1273			
Dd	68	TCAAGGTGATGTCATCATTGTGTAGAGATGCGTGGGTGAGAAAGTATGTTGGAGACGTTGG	127			
OY	1274	ATTTGGCAAAGCATATTACAAAGCAATGACCAAAATTCATAATTAATTAATTTTAAAGGA	1333			
Dd	128	ATTTGCGCCAAGACATATTACAAAGCTCTCTCAACTTCAGTGA AAAAACATTTTCAATGGTA	187			
OY	1334	ATGAGCATTTGCAAGTATGAAACAATTTGTAAACGCTTCATGTTCCCTTGGACAGGAAGCTA	1393			
Dd	188	ATGGTGTCATTGCTGTACATGGAACATTTGCAATGATTTATGTGGTGGAACTGGAAGCCA	247			
OY	1394	TCTCTTTGGTGTGTGTGTGTATGACTTTTGTGTGACAGGACCCCTGTGGTATCCAAAGC	1453			
Dd	248	TATCCCTTGGT	307			
OY	1454	GTAGCTTTTGGCTGCCAAGATGTCATCAGTCTCATATTTGGTCCCAAGCACGAGCTTGGATGG	1513			
Dd	308	GTACATATTGGCTACAAAGATGTCACTATGTGTGTCATTTGTGCTATCAATATGCTATGATGG	367			
OY	1514	GAACCTTCATCCACCCTGACGTGGATATGTTCACATCCACCCACCCCTTGTGGCGGCTTCC	1573			
Dd	368	GAATTTTATTCACCAACCAAGATTTGGGATATGTTTCAATATCAACACCCCTTGTGGCTTTTC	427			
OY	1574	ATGCTGCCTCTCGAGCCATCTGTGTGTGGCCGATCTATGTTAGTATTTCTGTGGCAAGC	1633			
Dd	428	ATGCAAGCTTTCAGAACCCATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	487			
OY	1634	ATAACTTTGATCTTGTGAAAACCTAGTAGCTTCCGATGGATGCAATCCCTGCAAGAGGT	1693			
Dd	488	ACAATTTTGACCTTCTCAAGAACTAGCTTGGCTGATGTGTACCAATCTTATATGTGAAA	547			
OY	1694	ACTATGCACTCCCGACCTCGCATTTGTTTGGATGGAAGCCCTTGGCATATGAGAACTA	1753			
Dd	548	CCTATGCTCTCCCAACCTAAGGACGTCTCTTTTTCGATCCCTTATACATGATGCAACAACTA	607			
OY	1754	TGCTTAGATTTTGGATCTCAACAAAGTTCACCTGAGTGTATGTGTGTCATTAACCTGCCAAG	1813			
Dd	608	TGCTCAAAATTTGGAACTCAATAAGTACACTGGAGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGT	667			
RESULT 2	A1727515	A1727515				
LOCUS	A1727515	646 bp	mRNA	linear	EST 11-JUN-1998	
DEFINITION	BNI6H18316 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (AB073744) raffinose synthase [Gossypium sativus], mRNA sequence.					
ACCESSION	A1727515					
VERSION	A1727515.1	GI:5046367				
KEYWORDS	EST.					
SOURCE	upland cotton.					

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Gossypium hirsutum	1	Blewitt, M., Matz, E.C., Davy, D.F. and Burr, B.	Ests from developing cotton fiber	Unpublished (1999)	Contact: Ben Burr
Eukaryotes: Viridiplantae: Streptophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots, Kossidae, eucotsids II: Malvales: Malvaceae: Gossypium.	1 (bases 1 to 646)				
Upton, NY 11973, USA					
Tel.: 516-344-3396					
Fax: 516-344-3407					
Email: burrb@nsl.bnl.gov					
Seq primer: T3 Primer					
Location/Qualifiers					
Source					
I. .646					

FEATURES	Seq primer: T3 Primer.
Source	Location/Qualifiers
	1. .646
	/organism="Gossypium hirsutum"
	/cultivar="Acala Maxxa"
	/db_xref="taxon:3635"
	/clone_lib="Six-day Cotton fiber"
	/lssue_type="Immature fiber"
	/dev_stage="Six days post anthesis"
	/lab_host="XLI-Blue"
	/note="Vector: pBluescript II KS+"
BASE COUNT	162 a 139 c 149 g 193 t 3 others
ORIGIN	
Query Match	17.1%; Score 402; DB 9; Length 646;
Best Local Similarity	76.3%; Pied. No. 1.4e-107;
Matches 492; Conservative	0; Mismatches 153; Indels 0; Gaps 0
OY 1083	GCAATGACCATGAGGATTTGGCGGTGGATAAGATTGTTCTTCAATAGTGGCGCTGT 1142
Db 1	GAAGAAAACGATGAGGATCTTCCGCCGTAAGATTGCAATATTGGTATCGATTAGT 60
OY 1143	CCCCCGAGAGAGCGTCGAGAGATGTACGAGAGACTTCATGCTCATTTGGAAAAAGTTGG 1202
Db 61	GCCCGCCGAGATGCGTCATCACTTTATGAGGAATTCATTTCTCATTTGGAAAAAGTTGG 120
OY 1203	GATCGACGCTGTATAGATTGACGCTTATCCACCTATTGGAGATGTTGTGAAGACTATGG 1262
Db 121	AATTGACGAGGATTAAAGTTGACCTGATTCACTGTGTGAAATGCTGTGAAAACCTATGG 180
OY 1263	AGGAGAGTGATTTGGCAAGGCAATATTACAAGCAATGACCACCAATCAATTAATAACA 1322
Db 181	TGGAAGGATTGATCTTTCACAAAGCTATTATTAAGCTTTAAACAGATTACAGTTAAAAAGCA 240
OY 1323	TTTTAAAGAAATGAGATCATTTGCAAGTATGAGACATTTTAACGACTTCATTTGCTTTGG 1382
Db 241	TTTCAAGGCAATGCTGTATTGTCCAGTATGGAACACTGCCAAGATTTCATGTTCTTTGG 300
OY 1383	CACGGAAGCTATCTCTCTTGGCTGTGTGGGTGATGACCTTTGGTCCAGGAGCCCTTGG 1442
Db 301	AACGAGACCATTTGTCTTGGTGTCTGGGACGATTTTGTGCTACGATTCATCAGG 360
OY 1443	TGATCCAAACGGTATCGTTTGGCTCCAAAGATCTCAGATGTTCAATTGTGCCAAGCAGAG 1502
Db 361	TGACCTTAATGGGCACTTTTGGCTCCAAAGGTGTCCAGTGGTGCACGTGTGCTTAACAAAG 420
OY 1503	CTTGTGATGGGGAACCTTCATCCACCCTGAGGGATGTATGTTCCATCCACCACACCTTGG 1562
Db 421	TTTATGATGGGCACTTTTATCCACCCTGATTGGGACATGTTCCAGTCTCTCCACACCTTG 480
OY 1563	TGCCGCTTCATGCTGCCCTCTGAGGACATCTCTGGTGGCCCGATCTATGTTAGTATTC 1622
Db 481	TGCTGATGTTCCAGCGTCTTCAAGGGCCATCTCCGGTGGCCCAATTTACATCANCAGAC 540
OY 1623	TGTGGGAACATTAACCTTTGATCTTTTGAAAAACTAGTGCCTCTCGTGAATGATGATCTT 1682
Db 541	TGTTGGCAACCACTTTGCTCTCTCTCAAAAGCGCTCGTNTTACCGAGTGGTGTGATCTT 600

BASE COUNT ORIGIN	177 a	133 c	144 g	162 t
Overall Match	15.48	Score 363.3	DN 10	Length 616

JOURNAL
COMMENT
Unpublished (2000)
Contact: May GD

OY	1519	TTCATCCACCGTCACTGGGATATAGTTCCAAATCACACCACCGCTTGCCGCCTTCATGCT	1578
Db	62	TT	121
OY	1579	GCCCTCAGACCATCTCTGTGTGGGCCGATCTATTGTAGTATCTGTGGGAAGAAGCATTAAC	1638
Db	122	GGCTCAAGAGCCATATCTGTGTGGGCCCATTTTTCATCATGACACATTTGGGAAGCCAACAC	181
OY	1639	TTTTGATCTTCTGAAAAAACAATACTAGTCTCTGATGATGCATCTTCGAAGTAGACTAT	1698
Db	182	TTTGGACCTCTTAAGACCTTTGGGCTCCAGATGGGCTCCATCTTCAGATGTGAGCACTAT	241
OY	1699	GCACATCCCAGCTCAGCATTTGTTGTTTGAAGACCCCTTTGCATTAATGAGAAACTATAGTT	1758
Db	242	GCACATCCCAACCAAGGACACTGTCTCTTGTGTGACCTCTTCATGATGAGCAAAAATGCTTC	301
OY	1759	AAGATTGGAAATCTCAACAAGTTCACTGGAGTAGATGGTATTCATTCACATGCCAAGGAGA	1818
Db	302	AAGATTGTGAAGCTTCAACAAGTAACTAGTAGATTTCTTGGGGTGTTTAAGTCCACAGGAGGA	361
OY	1819	GGATGGTGTCTGAGACACAGCCGCAACAATGCTTTTTCACAATFACTCAAAACGAGTAGACA	1878
Db	362	GGTGTGTTCCGAGAGTATGAGTTCACAAATATGCTGTGAGATTTTCTATAGGGATATCA	421
OY	1879	TCCAAACATCAACCAAAAGACATAGATAGGACAGTGGAGAAACCCCTATCTATTTGAA	1938
Db	422	ACCAAAGCCAAATATCAAAAGACATGATGATGGATGATGGAAGAATCCAATTTTCATTTGAA	481
OY	1939	GGCCTTTAAACCTTTGGCGCTTACCTCTATCAACGCCAAAAAATTATCCTCTCCACGCC	1998
Db	482	GGGGTGCAACTTTTGCTTCGTATTTACCAAGCAAGCAAGAACAATCTATCTCAGACACA	541
OY	1999	TCTCAAGATCTGA	2012
Db	542	TCTGATGACAGTGA	555

RESULT 8
BF269676

LOCUS GA_EB0005F08f Gossypium arboreum 7-10 dpa fiber library Gossypium

DEFINITION GA_EB0005F08f Gossypium arboreum 7-10 dpa fiber library Gossypium

ACCESSION BF269676

VERSION BF269676.1 GI:11200671

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

Enkayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

REFERENCE 1 (bases 1 to 564)

AUTHORS Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.

An integrated analysis of the genetics, development, and evolution of the cotton fiber

Unpublished (2000)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAAATCGACTACTATAGG

High quality sequence stop: 563.

location/Qualifiers

1..564

/organism="Gossypium arboreum"

/strain="AKA"

/cultivar="8400"

/db_xref="taxon:29729"

/clone="GA_EB0005F08f"

FEATURES

source

[illegible]

Db	244	CCTGAGATGTGAGCACTATGCATCCCATGAAGACATGTCTTTTGGTGAATCTCTCCA	303
Oy	1740	TAAATGAGAAACTATGCTTAAGATTTGGAAATCTCAACAAGTTCACTGGAGTATTTGGTC	1799
Db	304	TGATGGCAAAACAAATGACTCAGATATGGAACTCAACATGACACTGGAGTTCTTGGGGT	363
Oy	1800	ATTTCACTGCCAAGGAGGAGTGTGTCTGTGAGACACGCCCAACCAATGCTTTTTCACA	1859
Db	364	ATTAACTACGCCAAGGAGGAGTGTGTCTGTGACATGATTCACAACAAAGGTGCTGTGA	423
Oy	1860	ATATCCAAACAGCAGATCATCCAAACAACTAACCACCAAAAGCATAGATAGGCAAGTGGAGA	1919
Db	424	GTTTTCATCTGTGATTCACACCAATTCAGCATCCCAAGCATTTGATGGTGAACAGTGTGA	483
Oy	1920	AAACC 1925	
Db	484	GAAACC 489	

RESULT 11	
BF636592	513 bp
LOCUS	mRNA
DEFINITION	linear
ACCESSION	EST 19-DEC-2000
VERSION	clone NF090H02D1F1027
KEYWORDS	5', mRNA sequence.
SOURCE	Barrel medic.
ORGANISM	Medicago truncatula

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 513)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J...
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
Unpublished (2000)
Contact: May GD

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdamay@noble.org
Insert Length: 513 Std Error: 0.00
Plate: 090 row: H column: 02
Seq primer: TCACACAGGAACAACGCTATGAC.

FEATURES	SOURCE	Location/Qualifiers
		1. . 513
		/organism="Medicago truncatula"
		/db_xref="taxon:3880"
		/clone="NF090H02PT"
		/clone_lib="Drought"
		/tissue_type="Plasmids"
		/dev_stage="Pooled timepoints"
		/note="Vector: Lambda Zap; Contains a mixture of entire plasmids harvested in a series of days-post-watering timepoints."
BASE COUNT		135 a 72 c 138 g 167 t 1 others
ORIGIN		

	Query Match	12.5%	Score 293.4	DB 12	Length 513
	Best Local Similarity	73.2%	Pred. No. 2,1e-75		
	Matches 375	Conservative 0	Mismatches 137	Indels 0	Gaps 0
OY	1000	TGTGATATTTGGGGTGGCCCTTCGCCGAGGTGCTGCTGCTGCTGAGGACGCTGTGATT			1059
Db	2	TGTGATATTTGGGGTGGGATTAAGACCTATATGTTGAGGGCTTACCTGAAGACTTATTTGGAG			61
OY	1060	CAGCCAGTGCCTTTCACCCAGGCGCTGAGATGACGATGAGGAGGATTTGCGCGTGATTAAGATT			1119

Db	62	AGCCTTAAGTACCCTTGATGGAAACTACTATGAGAGATTGGCTGTGATTAAGATT	121
Qy	1120	CTTCTTCATTAAGTCGGGCTGCTCCCGCGAGAAGCTGAGAGATTAGCAGAGCTT	1179
Db	122	GTTAAATAATGCTGTGGGTAGTCTCTCTCATTTGGTTGATCAATATGATGAAGTATT	181
Qy	1180	CATGCTCATTTGGAAAAAGTTGGATTCGACGCTGTTAAGATTGACGTTATCCACTATTG	1239
Db	182	CACCTCTCTTTGGAAAGAGCTGGAAATTGATGGGCTCAAGTGTACATCATTTGGCTA	241
Qy	1240	GAGATGTGTGGAAGACTATGAGAGGAGAGTGGATTTGGCAAAGCATATTACAAGCA	1299
Db	242	GGAATGCGGGAGCAAGATATGTTGGAAAGATTGATTTGGCCAAAGCATATTACAAACT	301
Qy	1300	ATGACCAATCAATTAATTAACATTTTAAAGAAATGGATTCATTGCCAAGTATGGAACAT	1359
Db	302	CTCTCAACTCTAGTAAAAAACATTTTCAATGGTAATGGTCTATTGCTAGCATGGAACAT	361
Qy	1360	TGTAAAGACTCATGTTCTCTGGACGAGAGTATCTCTCTTGCTGCTGTGGATGAC	1419
Db	362	TGCATGATTTATGTGTGGAACTAGAGCATATCCCTTGGCTGCTGTGGATGAT	421
Qy	1420	TTTTGGTCAGCAGCCTCTGTGTGATCCAAACGGTACGTTTGGCTCCAAGATGTAC	1479
Db	422	TTCGTGTGACAGCCCATATGCTGATCCAAATGAGACATATTGGCTCAAGAGATGCAC	481
Qy	1480	ATGTTTCATTGTCCACAGCAGCTTTGGAT	1511
Db	482	ATGTGCATTTGCAATCAATATCTATTGAGAT	513

RESULT	12
BG321136	
LOCUS	BG321136 818 bp mRNA linear EST 27-FEB-2001
DEFINITION	zm04_05d11-R zm04_AAF-C_ECORC_cold_stressed_maize_seedlings Zea mays
ACCESSION	BG321136
VERSION	BG321136.1 GI:13150814
KEYWORDS	EST.
SOURCE	Zea mays.
ORGANISM	Zea mays

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta Spermatophytas; Magnoliophytas; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.	Singh,J.A., Wakui,K., Couroux,P., De Moors,A., Harris,L.J., Hattori,T. J.,I., Oneilset,T., Robert,L.S., Sprolt,D. and Tinker,N.A.	Expressed Sequence Tags from Cold-Stressed Maize Seedlings Unpublished (2001)	Contact: Singh,J.A.	Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada KM Neelby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A 0C6, Canada 759-1662 Tel.: (613) Fax: (613) 759-1701 Email: singhja@em.agr.ca.
FEATURES	Location/Qualifiers	Source	1.	.818

```

clone.lib="Zm04.AAF.ECORG_cold_stressed_maize_seedlings"
/tissue.type="leaf, crown"
/notes="Vector: Bluescript SK-/XhoI-ECOR1. Site_1: Eco RI,
Site_2: Xho I. Lower temperature 50 c / hour from 22 to
120c; bring to 50 in 1 hour from 120c. Leave at 50c 2 days
, photoperiod 16 hours. Light intensity was 125 ue-1.
Library prepared by in vivo mass excision from amplified
library."
BASE COUNT      112 a      300 c      230 g      129 t      47 others

```


QY	1786	GGAGTATGTGGGTCATTAACATGCCAAGAGGAGATGGTGTGCGAGACAGCCGCAAC	1845
QY	1786	GGAGTATGTGGGTCATTAACATGCCAAGAGGAGATGGTGTGCGAGACAGCCGCAAC	1845
Db	241	GGAGTCTTGCGGCTTTAACTGAACGCCAGGAGAGGTGGTGGCGTGAAGTTAGTGTCCAAC	300
QY	1846	CATGCTCTTTTACCACTACTCCAAAAGAGTGCATCCAAAATCTAACCCAAAGACATAGAA	1905
Db	301	AAATGTGCTGCGAATTTTCTCATAGGGATATCAACCAACCCAAATATCAAAACATTTGAA	360
QY	1906	TGGCAGTGTGAGAAACCCATCTCTATGTGAAGCGGTAAACCTTTGCGCTTAACCTC	1965
Db	361	TGGATTAAGTGGAAABAAATCCATTTTCCATTTGAAGGGGTCCAACTTTTGGCTTCGTATTTC	420
QY	1966	TATCAAGCCAAAACCTTATCTCTTCCAAAGCCCTCTCAAGA--TCCTGATTAAGCTCTT	2022
Db	421	AGCCCAAGCCAGAAACCTATCTCTCAGACACCATCTGATGACAGTAGAAGATTTCCCTTG	480
QY	2023	GACCATTGCGAATGGAGCTCATCTGTTTATACCAAGTAGC 2063	
Db	481	GAGCCATTCAATTTGAGCGTATATAACAGATTTTCCCTGTGAC 521	

RESULT 14	LOCUS	DEFINITION	LOCUS	DEFINITION
BF425566	BF425566	525 bp	mRNA	linear
LOCUS	LOCUS	LOCUS	LOCUS	LOCUS
DEFINITION	DEFINITION	DEFINITION	DEFINITION	DEFINITION
Gm-c1068-2061 5'	Gm-c1068-2061 5'	similar to TR:Q9ZT62	Q9ZT62	RAFFINOSE SYNTHASE ;
mRNA-sequence.	mRNA-sequence.	mRNA-sequence.	mRNA-sequence.	mRNA-sequence.

ACCESSION	BF425566	GI:11413555
VERSION	BF425566.1	
KEYWORDS	EST.	
SOURCE	soybean.	
ORGANISM	Glycine max	

ORGANISM	REFERENCE
Glycine max	1 (bases 1 to 525)
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: tracheophyta:	
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:	
Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Phaseoleae:	
Glycine:	
Shoemaker, R., Ksilm, P., Vockin, L., Erpelding, J., Coryell, V., Khanna	

REFERENCE
AUTHORS
1 (bases 1 to 525)
Shoemaker, R., Kelm, P., Vockin, L., Erpedding, J., Coryell, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Y., Pearson, B., Swaller, T., Gibbons, M., Pape, D., Haren, N., Bowers
R., Ritter, B., Kohn, S., Sulin, T., Jackson, Y., Cardenas, M.,
R., Walter, R., and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

TITLE	JOURNAL	COMMENT
Public Soybean EST Project	Unpublished (1999)	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA tel: 314 266 1800

Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: cu@resgen.com
Insert length: 1160 Std Error: 0.00
High quality sequence stop: 320.

FEATURES
source

```

1. .525
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1068-2061"
/clone_id="Gm-c1068"
/tissue_type="Leaf, drought stressed, 1 month old plants,
greenhouse grown"
/lab_host="DH108"
/notes="vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from drought stressed leaf tissue of the cultivar Williams
82. The month old greenhouse grown plants were deprived of
water for 3 days prior to harvesting the stressed leaf
tissue. Complementary DNA was synthesized from mRNA using
primer consisting of a poly(dT) sequence with a XhoI

```

BASE COUNT 132 a 132 c 118 g 139 t 4 others

ORIGIN

Query Match	11.4%	Score 268:	DB 12:	Length 525:
Best Local Similarity	69.4%	Pred. No. 7.4e-68:		
Matches 361; Conservative	0:	Mismatches 159:	Indels 0:	Gaps 0

QY 1487 ATTGTCACACGACGCTTGTGGATGGGAACTTCATCCACCCTGACTGGGATATGTTCC 1546

Db 2 ATGTGCATACACAGCTTGTGGATGGGCAATTTCATCCACCACAGATTGGGACATGTCC 61

QY 1547 AATCCACCACCTGTGCGCCTTCATGCTGCCCTCGAGCCATCTCTGGTGCCCGA 1606

Db 62 AATCTACTATCCTTGTGCTGCTTCATGCTGCCCTCAAGAGCCATATCTGGTGCCCCA 121

QY 1607 TCTATGTTAGTGATTCGTGGGAAGCATAACTTTGATCTTCTGAAAAAACTAGTGCTTC 1666

Db 122 TTTACATCAGTGACACAGTTGGGAACCACTTTGAGCTGCTTAAGACCTTGGCCTTGC 181

QY 1667 CTGATGGATCGATCCTTCGAAGTGAGTACTATGCACCTCCGACTCGGATTGTTGTTG 1726

Db 182 CAGATGGTCCATCCTCAGATGTGAGCACTATGCACCTCCCAACGAGGACTGTCTCTTG 241

QY 1727 AAGACCCCTTTCATTAATGGAGAACTATGCTTAAAGATTGGAATCTCAACAAGTTCACCTG 1786

Db 242 CTGACCCCTCTCCATGATGGCAAAACAATGCTCAAGATATGGAACCTCAACAAGTACACTG 301

QY 1787 GAGTGATTGGTCATTCAACTGCCAAGGAGGAGGATGGTGTCTGTGAGACACGCCCAACC 1846

Db 302 GAGTTCCTTNGGTGTTAACTGCCAGGAGGAGGTTGGTTCCTCCNTGAGATTANGTCCANCA 361

QY 1847 AATGCTTTTCAACAATACTCAAAACGAGTGACATCCAAACTAACCCTAAAGACATAGAAT 1906

Db 362 AATGTGCTGCTGAGTTTCTCATCGGTGATCATCCAGAACCATATGAAAGACATTGAAT 421

1907 GGCACAGTGGAGAAACCTATCTCTATTGAAGCGTTAAACCTTTGGCGTTACCTCT 1966

Db 422 GGATAGTCGAACAATTCATGTACCATTTGGAGGGTGCACCTTTTCTCGTACGTATGCA 481

QY 1967 ATCAAGCCAAAACCTTATCCTCTCCAGCCCTCTCAAGA 2006

Db 482 ACCAAGCCCAAGAACTCATTCCTCTCGGGGCCATTGATGA 521

RESULT 15

AT993310	530 bp	mpna	linear	EST	08-SEP-
LOCUS	AT993310				

DEFINITION	701496022 A. thaliana, Ohio State clone set	Arabidopsis thaliana
CDNA clone	701496023	mpna sequence

ACCESSION	AI993310
VERSION	AI993310 1
	GT:5840215

KEYWORDS EST. +ha]e creese SOURCE

ORGANISM *Arabidopsis thaliana*
Eukaryota: Viridiplantae: Streptophyta: Tracheophyta

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Posidaceae; eucosids II; Brassicales; Brassicaceae; Arabidopsis

REFERENCE	1 (bases 1 to 530)
AUTHORS	Chen T, Momiyama M, Chan F, Moonav M, Carreon B, Cillitland

Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Corzone, C., Burns, D., Griffin, T., Mouanoutoua, M., Neuen, D., Ta

Pollock, T. Suzuki C. Argentine C. Shah S. Nobrega A. Murray
, Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T.

Turner, C., Krikorian, S., Elder, L. and Hanson, D.
Arabidopsis thaliana Gene Expression Microarray

JOURNAL Unpublished (1999)

COMMENT

Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

FEATURES

Location/Qualifiers
1. .530
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone_id="A. thaliana, Ohio State clone set"
/note="CDNA library was made from selected clones from the
Arabidopsis thaliana Ohio State clone set."

BASE COUNT 132 a 115 c 128 g 155 t

ORIGIN

Query Match 11.3%; Score 266.6; DB 9; Length 530;
Best Local Similarity 69.0%; Pred. No. 1.9e-67;
Matches 365; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

```
OY 1613 TTACGATTCCTGTGGGAAGCATATCTTGATCTTGAAAAACTAGTCTCTCTGATG 1672
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 TCAGGATTTGTGTGGCAAGCATGATTTGATCTTGGAAGCGCTGTTTGGCCACG 61
OY 1673 GATCATCTTCGAGTAGTACTATGCACTCCGACTGCGATTTGTTGTTGAAGAC 1732
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 GTTCGATTTTGAGTGTGAGTACTGCTCTGCTCCCACTGTCACCGTCTCTTTGAGATC 121
OY 1733 CTTTCATTAATGAGAACTATGCTTAAGATTGGAATCTCAACAAGTTCACTGAGTGA 1792
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 122 CTCCTCATGATGGCAAAACGATCTCAAGATTGGAACCTGGAACAAGTACACTGGAGTTA 181
OY 1793 TTGGTGCAATTCACAGCCCAAGAGAGATGCTGTCGTGAGACAGCCGCAACCAATGCT 1852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 182 TTGGAGCATTCACAGTGTCAAGAGAGAGATGGTGCAGAGAAACAGACGTAAACCAATGTT 241
OY 1853 TTTCACAATTAACCAAGAGAGATGATCCAAACTAACCAAAAGACATAGAAATGGCACA 1912
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 TCTCTGATTCCTCAACACGTTAACCGCCACCAACAGCCCTTAAGACGTTGAATGGAACA 301
OY 1913 GTGAGAAAACCTTATCTCTATTTGAAGGCGTTAAAACTTTGCGCTTACCTTATCAAG 1972
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 GTGGAAGCAGCCCAATCTTCATTCGAAACGTTGAAGAGTTTCTTGTCTCTCAAT 361
OY 1973 CCAAAAACCTTCTCTCTCCCAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTCG 2032
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 CCAAGAGAGCTTTGTTGCTGTGAGACTAAACGATGATCTTGAGCTTGAGACCTTTCA 421
OY 2033 AATTGAGCTCATCACTGTTTCCAGAGTGAACCAACTCATCCAAACTTCTTACACTTTG 2092
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 AGTTGAGTGTATCACTGCTCTCTCTGTTGTGACCATGAGGTAATTCAGTCGGTTTG 481
OY 2093 CCCCAATTGGGCTGTGGAACATGCTTAACACTAGTGAAGCCATCCAATC 2141
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 482 CTCGATTTGAGTGGTAAACATGCTAAACACAAAGCGGTGCGATCCGGTC 530
```

Search completed: April 3, 2003, 19:22:25
Job time : 3798 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 16:07:45 : Search time 100 Seconds
(without alignments)
7213.037 Million cell updates/sec

Title: US-09-675-208-4_COPY_56_2407
2352
Sequence: 1 ATGGCTCTAGATTAA...CGGTATGAGTACTGTTT 2352

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/6C.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2352	100.0	2517	4	US-08-846-234-4
2	976.4	41.5	2498	4	US-09-561-825-23
3	47.4	2.0	7218	1	US-08-232-463-14
4	38	1.6	911	4	US-09-457-046B-19
5	38	1.6	911	4	US-09-457-046B-21
6	36.8	1.6	19736	4	US-09-740-035-3
7	36	1.5	2824	2	US-09-010-928B-3
8	36	1.5	8083	4	US-09-383-630-4
9	36	1.5	8083	4	US-09-383-630-5
10	35.6	1.5	2793	1	US-08-209-747-1
11	35.6	1.5	2793	1	US-08-458-298-1
12	35.2	1.5	2338	1	US-08-425-069-1
13	35.2	1.5	2338	2	US-08-317-844B-1
14	34.8	1.5	1326	4	US-09-457-046B-53
15	34.8	1.5	18596	4	US-09-318-448-11
16	34.6	1.5	289	4	US-09-007-005-17
17	34.6	1.5	289	4	US-09-244-796-17
18	34.4	1.5	216	1	US-07-609-716-45
19	34.4	1.5	216	3	US-08-475-411A-45
20	34.4	1.5	216	4	US-08-478-029A-45
21	33.8	1.4	7218	1	US-08-232-463-14
22	33.6	1.4	1416	3	US-08-911-853-3
23	33.6	1.4	1416	4	US-09-479-409-3
24	33.6	1.4	1416	4	US-09-479-453-3
25	33.6	1.4	2509	2	US-08-014-969-1
26	33.6	1.4	4377	3	US-08-911-853-28
27	33.6	1.4	4377	4	US-09-479-409-28

C 28	33.6	1.4	4377	4	US-09-479-453-28	Sequence 28, Appl
C 29	33.6	1.4	6234	4	US-08-971-188-1	Sequence 1, Appl
C 30	33.4	1.4	1505	1	US-07-915-246-1	Sequence 1, Appl
C 31	33.4	1.4	16442	3	US-08-781-891-208	Sequence 208, App
C 32	33	1.4	2561	4	US-09-270-542-101	Sequence 101, App
C 33	33	1.4	2561	4	US-09-270-542-119	Sequence 119, App
C 34	33	1.4	34303	2	US-08-735-609-4	Sequence 4, Appl
C 35	33	1.4	34303	4	US-08-735-609-4	Sequence 4, Appl
C 36	33	1.4	34303	3	US-09-315-372-4	Sequence 4, Appl
C 37	33	1.4	34303	3	US-09-244-752-4	Sequence 4, Appl
C 38	33	1.4	34303	3	US-09-245-497-4	Sequence 4, Appl
C 39	33	1.4	34303	4	US-09-562-919-4	Sequence 4, Appl
C 40	33	1.4	34382	2	US-08-374-483-6	Sequence 6, Appl
C 41	33	1.4	35408	4	US-08-973-334-3	Sequence 3, Appl
C 42	33	1.4	35408	4	US-09-563-869A-3	Sequence 3, Appl
C 43	33	1.4	35408	4	US-08-549-489-3	Sequence 3, Appl
C 44	33	1.4	35935	2	US-08-735-609-1	Sequence 1, Appl
C 45	33	1.4	35935	2	US-08-735-609-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-846-234-4
Sequence 4, Application US/08846234
Patent No. 6166292
GENERAL INFORMATION:
APPLICANT: OSUMI Chieko
APPLICANT: NOZAKI Jinshi
TITLE OF INVENTION: KIDA Takao
TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA: US/08/846, 234
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-413-3000
TELEFAX: (703)-413-2220
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2517 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: cucumber (Cucumis sativus)
FEATURE:
NAME/KEY: CDS
LOCATION: 56..2407
US-08-846-234-4
Query Match 100.0% Score 2352; DB 4; Length 2517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCTAGTTTAAAAATGSGCTCCACAGTAGTTTCATTGATGAGCTTAAATGAC 60
|||||
Db 56 ATGGCTCTAGTTTAAAAATGSGCTCCACAGTAGTTTCATTGATGAGCTTAAATGAC 115
QY 61 ATGCTGTACCGTTTGCAATCGACGATCGGATTTCACTGTGAAGGTCATTCGTTTCTG 120
|||||
Db 116 ATGCTGTACCGTTTGCAATCGACGATCGGATTTCACTGTGAAGGTCATTCGTTTCTG 175
QY 121 TCCGAGTTCCTGAGAAACATGTTGCTTCCTTCCTCCGACACTTCGATGAGCAAGTCC 180
|||||
Db 176 TCCGAGTTCCTGAGAAACATGTTGCTTCCTTCCTCCGACACTTCGATGAGCAAGTCC 235
QY 181 CCGGTTTCGGTGGTTCCTTTGTTGATTCGACGCGTCGGAACTGATAGCCGACATGTT 240
|||||
Db 236 CCGGTTTCGGTGGTTCCTTTGTTGATTCGACGCGTCGGAACTGATAGCCGACATGTT 295
QY 241 GTTTCATTTGGAGAGCTGAAGATATTCGTTATGAGTATTTTCAGGTTTAAAGCTTTGG 300
|||||
Db 296 GTTTCATTTGGAGAGCTGAAGATATTCGTTATGAGTATTTTCAGGTTTAAAGCTTTGG 355
QY 301 TGGACTACACACTGGGTGGTCGAATGGTGGGATCTTGAATCGGAGACTCGATTTGTG 360
|||||
Db 356 TGGACTACACACTGGGTGGTCGAATGGTGGGATCTTGAATCGGAGACTCGATTTGTG 415
QY 361 ATCCCTTGAGAAAGTCAGATTCGTGTCACCGTATGTTTCCCTTCCTCGATCGTTGAAGGA 420
|||||
Db 416 ATCCCTTGAGAAAGTCAGATTCGTGTCACCGTATGTTTCCCTTCCTCGATCGTTGAAGGA 475
QY 421 CCGTTCGGAACCTCGATTCAGCTGGGAGATGATGACTTTTGATGATTTTGTGTGAGAGT 480
|||||
Db 476 CCGTTCGGAACCTCGATTCAGCTGGGAGATGATGACTTTTGATGATTTTGTGTGAGAGT 535
QY 481 GGTTCCTGCAAAAGTTTGTGATGATGATCGATCGATCGATGATGATGATGATGATGAT 540
|||||
Db 536 GGTTCCTGCAAAAGTTTGTGATGATGATCGATCGATCGATGATGATGATGATGATGAT 595
QY 541 GATCCGTTTGACACTGTTTAAAGAGCGGATGAGATCGTGTGAGAACCCATCTTGGAACTTTT 600
|||||
Db 596 GATCCGTTTGACACTGTTTAAAGAGCGGATGAGATCGTGTGAGAACCCATCTTGGAACTTTT 655
QY 601 CGCTTTGGAGAGAGAACTCCACAGGATGATGAGCAAAATTCGGTTGATGATGATGATG 660
|||||
Db 656 CGCTTTGGAGAGAGAACTCCACAGGATGATGAGCAAAATTCGGTTGATGATGATGATG 715
QY 661 GAGCGCTTTTACCTAAGCGTTTCATCCACAGGCGGTATAGAAAGCGGTGAAGCATCTCGTC 720
|||||
Db 716 GAGCGCTTTTACCTAAGCGTTTCATCCACAGGCGGTATAGAAAGCGGTGAAGCATCTCGTC 775
QY 721 GAGCGCGGTTTCTCCCGGTTTATGTCCTAATCGACGATGCTGGCAATCCATCGAGCAC 780
|||||
Db 776 GAGCGCGGTTTCTCCCGGTTTATGTCCTAATCGACGATGCTGGCAATCCATCGAGCAC 835
QY 781 GATTGGATCCCATCACAAAGAAATGAACCAACCGTCCGCGGCGCAAAATGCCCC 840
|||||
Db 836 GATTGGATCCCATCACAAAGAAATGAACCAACCGTCCGCGGCGCAAAATGCCCC 895
QY 841 TGGCGCTTTTGAATTTCCAAAGAAATTAACAATTCCTGTGACTACTGCAATCCCAAGGCC 900
|||||
Db 896 TGGCGCTTTTGAATTTCCAAAGAAATTAACAATTCCTGTGACTACTGCAATCCCAAGGCC 955
QY 901 ACCGCGCCCCGAGCGCGCCAGAAAGGAGTGAAGCGCTTATAGATGAACCTCAAAAGAGAG 960
|||||
Db 956 ACCGCGCCCCGAGCGCGCCAGAAAGGAGTGAAGCGCTTATAGATGAACCTCAAAAGAGAG 1015
QY 961 TTTTAAAGCTGTGAGCATGTTTATGTTTGGCATGCTTTGTGTGGATGATTTGGGTGGCCTT 1020
|||||
Db 1016 TTTTAAAGCTGTGAGCATGTTTATGTTTGGCATGCTTTGTGTGGATGATTTGGGTGGCCTT 1075
QY 1021 CGCGCCGAGGTGCTGCTTGCCTGTGAGGCACGTGATTTACAGCCAGTGTCTTACCAGAGG 1080
|||||
Db 1076 CGCGCCGAGGTGCTGCTTGCCTGTGAGGCACGTGATTTACAGCCAGTGTCTTACCAGAGG 1135
QY 1081 CTGACAGATGACATGAGAGATTGTCGCGTGTGATTAAGATTGTTCTTCATAAAGTTCGGGCTG 1140

Db 1136 CTGACAGATGACATGAGAGATTGGCGGTGATGAATGTTCTTCATAAGCTGCGGCTG 1195
|||||
QY 1141 GTCCCGCCGAGAGAGGCTGAGGAGATGTACGAAGGACTTCATCTGATTTGGAAAAAGTT 1200
|||||
Db 1196 GTCCCGCCGAGAGAGGCTGAGGAGATGTACGAAGGACTTCATCTGATTTGGAAAAAGTT 1255
QY 1201 GGGATCGACGGTGTAAAGATTGACGTTATCCACTTATGAGATGTTGTGTGAAGACTAT 1260
|||||
Db 1256 GGGATCGACGGTGTAAAGATTGACGTTATCCACTTATGAGATGTTGTGTGAAGACTAT 1315
QY 1261 GGAAGGAGAGTGAATTTGGCAAGCATATTTACAAAGCAATGACCAATCAATTAATTA 1320
|||||
Db 1316 GGAAGGAGAGTGAATTTGGCAAGCATATTTACAAAGCAATGACCAATCAATTAATTA 1375
QY 1321 CATTTTAAAGAAATGAGACTTGAAGATGAGCAATGATGATGATGATGATGATGATG 1380
|||||
Db 1376 CATTTTAAAGAAATGAGACTTGAAGATGAGCAATGATGATGATGATGATGATGATG 1435
QY 1381 GGCACGAAAGCTATCTCTCTGTCGTGTGTGATGATGATGATGATGATGATGATGATG 1440
|||||
Db 1436 GGCACGAAAGCTATCTCTCTGTCGTGTGTGATGATGATGATGATGATGATGATGATG 1495
QY 1441 GGTGATCCAAAGCGTACGTTTGGCTCCAAAGATGTACATGTTGATGTTGCCAAAGAC 1500
|||||
Db 1496 GGTGATCCAAAGCGTACGTTTGGCTCCAAAGATGTACATGTTGATGTTGCCAAAGAC 1555
QY 1501 AGCTTTGATGAGGAGAACTTATCCACCCGATGAGGATGATGATGATGATGATGATGATG 1560
|||||
Db 1556 AGCTTTGATGAGGAGAACTTATCCACCCGATGAGGATGATGATGATGATGATGATGATG 1615
QY 1561 TGTGCCGCTTTCATGCTGCTGCTGTCGAGGCATCTGTTGGGCCGATCTATGTTAGTAT 1620
|||||
Db 1616 TGTGCCGCTTTCATGCTGCTGCTGTCGAGGCATCTGTTGGGCCGATCTATGTTAGTAT 1675
QY 1621 TCTGTGGGAAGCATTAATCTTGTATCTTGTGAATAAACTAGTGTCTGATGATGATGATG 1680
|||||
Db 1676 TCTGTGGGAAGCATTAATCTTGTATCTTGTGAATAAACTAGTGTCTGATGATGATG 1735
QY 1681 CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740
|||||
Db 1736 CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1795
QY 1741 AATGAGAAACTATGCTTAAAGATTGGAATCTGACAAAGTTCATGAGTATGATGATG 1800
|||||
Db 1796 AATGAGAAACTATGCTTAAAGATTGGAATCTGACAAAGTTCATGAGTATGATGATGATG 1855
QY 1801 TTTCATCTGCCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860
|||||
Db 1856 TTTCATCTGCCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1915
QY 1861 TACTCAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920
|||||
Db 1916 TACTCAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1975
QY 1921 AACCTATCTCTATTTGAAGGCGTTTAAACCTTGGCCTTTACCTGATATCAAGCCAAAA 1980
|||||
Db 1976 AACCTATCTCTATTTGAAGGCGTTTAAACCTTGGCCTTTACCTGATATCAAGCCAAAA 2035
QY 1981 CTATATCTCTCCAAAGCGCTCTCAAGATCTTGAATAGCTTGAACCATTCGAATTCGAG 2040
|||||
Db 2036 CTATATCTCTCCAAAGCGCTCTCAAGATCTTGAATAGCTTGAACCATTCGAATTCGAG 2095
QY 2041 CTATATCTCTTACCAAGAGGACCAAACTCATCAAACTCTCTATACCTTTGCCCAATT 2100
|||||
Db 2096 CTATATCTCTTACCAAGAGGACCAAACTCATCAAACTCTCTATACCTTTGCCCAATT 2155
QY 2101 GGGCTGTGTAACATGCTTTAACTAGTGAAGCATCCAAATCTGTGATGATGATGATGATG 2160
|||||
Db 2156 GGGCTGTGTAACATGCTTTAACTAGTGAAGCATCCAAATCTGTGATGATGATGATGATGATG 2215
QY 2161 CTTAAGTCACTGAGATTTGTTGCAAAAGGTTGTGTGAAGATGAGATATTTGATGATGATA 2220
|||||


```
Db 1603 CATGTTCCAGTCACCTGCTGTGCGCAATTCATGAGCCTTAGGGCCATCTGTG 1662
Qy 1599 TGGCCCGATCTATTAGTACTGCTGGGAAGACATATCTTGTCTGTGAAAAACT 1658
Db 1663 TGGACCACTTACGTTAGTATGTTGTTGGAAAGCACAACCTTCAAGTCTCAAGAGCCT 1722
Qy 1659 AGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1718
Db 1723 CCTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1782
Qy 1719 TTTGTTTGAAGACCTTTCATATGAGAACTATGCTTAAGATTGGAATCTGACAA 1778
Db 1783 TTTGTTTGAAGACCTTTCATATGAGAACTATGCTTAAGATTGGAATCTGACAA 1842
Qy 1779 GTTCACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1838
Db 1843 ATATACAGGTGTTTGGTCTATTAATTTGCAAGAGAGTGGTGTGCTCCGTAAGT 1902
Qy 1839 CCGCAACCAATCTTTTCAATACTCAAAAGAGTACATCCAAAACCTAACCCAAAGA 1898
Db 1903 GAGAAACAAGAGTGCCTGATGATTTTCAAACTGTGATGATGATGATGATGATGATG 1962
Qy 1899 CATAGATGAGCAGTGGGAAAGACCTATCTCTATGAGGCGTTAAACCTTTGCGCT 1958
Db 1963 CATTAATGAGCAATGAGGAAAGACCAATATGCAATAAAGGATGATGATGATGATG 2022
Qy 1959 TTACCTCTATCAAGCAAAACTATCTCTCAAGCCCTCAAGATCTTGACATAGC 2018
Db 2023 ATATTGTTCAAGAGCACAACCACTAAAGCTCATGAAAGCATGAGAAATTTGGAATTC 2082
Qy 2019 TCTTGACCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 2078
Db 2083 ACTTACGATTCATTTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2142
Qy 2079 TTTCTACACTTGGCCCAATGAGGCTGCTGAGCAATGCTTAACATAGTGGAGCCATCA 2138
Db 2143 GTTAATTCATTTGCTCCCATTTGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2202
Qy 2139 ATCTTGACATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 2198
Db 2203 GTCATGAGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 2262
Qy 2199 GATGCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2258
Db 2263 GATGAGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2322
Qy 2259 GTTCAAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2318
Db 2323 ATTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 2379
Qy 2319 ATCG 2322
Db 2380 ATTG 2383
```

RESULT 3
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367

GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Hardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299

```
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)856-9300  
TELEFAX: (703)683-4109  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgt-F15  
US-08-232-463-14
```

Query Match 2.0%; Score 47.4; DB 1; Length 7218;
Best Local Similarity 4.7%; Pred. No. 0.00024;
Matches 15; Conservative 178; Mismatches 124; Indels 0; Gaps 0;

```
Qy 1077 AGGCGTCGATGAGCATGAGGATTTGGCGGTGATAGATTGTTCTTCATAGAGTGG 1136
Db 1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297
Qy 1137 GCTGTCGCCGAGAGAGATGAGAGATGACGATTCATGCTCATTTGGAAAA 1196
Db 1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237
Qy 1197 AGTTGGATCAGCGTGTAACTTGCCTTATCCACTATTGAGATTTGTGTGAGA 1256
Db 1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177
Qy 1257 CTAATGAGGAGAGTGGATTTGGCAAGCATATTACAAGCATGACCAATCAATAA 1316
Db 1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117
Qy 1317 TAAACATTTTAAAGAAATGAGATTCATTCAGATGAGAACTTGAACACTTCATGTT 1376
Db 1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1057
Qy 1377 CCTTGGCAGCAAGCTA 1393
Db 1056 CCTGACCTGCAGCCA 1040
```

RESULT 4
US-09-457-046B-19
Sequence 19, Application US/09457046B
Patent No. 6287835
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney et al.
TITLE OF INVENTION: Transcylases of the Facilitaxel Biosynthetic Pathway
FILE REFERENCE: 53679
CURRENT APPLICATION NUMBER: US/09/457,046B
CURRENT FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 19
LENGTH: 911
TYPE: DNA
ORGANISM: Taxus cuspidata
US-09-457-046B-19

Query Match 1.6%; Score 38; DB 4; Length 911;
Best Local Similarity 47.2%; Pred. No. 0.065;
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

OY 317 TTGGTCGCAATGGTGGGATCTTGAATCGGACACTGACATTGTGCTTCCTTGAGAACTCAG 376
DB 461 TTGTTATTAATGGTGCACACCATAGATATATGAAACAATGGTGTGATGAGAAATTAATG 520
OY 377 ATTCTGTGACCGGATGTTTCTTCTTCGATGCTTGAGGACCGTTCCGAACTCGA 436
DB 521 ATTTTGTTCGTCCTTTGAGTACGAGCAGCATTTGTTGGATGACCAAGACAAAGGCTC 580
OY 437 TTCACCGTGGGATGATGATCTTGTGATGTTGTGCGAGAGTGTTCGTAAGTTG 496
DB 581 TTCAATTCACATACATGAGATGGAAGCTTCTCTTGGATGATTTGAGAAATTAAT 640
OY 497 TTGATGACATCGTTCGGAATGTTGTATCTTCATGCTGGTATGATCCGTTGCACCTTG 556
DB 641 TTAATCCCCCACTTCCAATGATATATGTAATGCCATTGGTACTGATGCAATGG 700
OY 557 TTAAG 562
DB 701 ATAATG 706

RESULT 5
US-09-457-046B-21
Sequence 21, Application US/09457046B
Patent No. 6287835
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney et al.
TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
FILE REFERENCE: 53679
CURRENT APPLICATION NUMBER: US/09/457,046B
CURRENT FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 21
LENGTH: 911
TYPE: DNA
ORGANISM: Taxus cuspidata
US-09-457-046B-21

Query Match 1.6%; Score 38; DB 4; Length 911;
Best Local Similarity 47.2%; Pred. No. 0.065;
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

OY 317 TTGGTCGCAATGGTGGGATCTTGAATCGGACACTGACATTGTGCTTCCTTGAGAACTCAG 376
DB 461 TTGTTATTAATGGTGCACACCATAGATATATGAAACAATGGTGTGATGAGAAATTAATG 520
OY 377 ATTCTGTGACCGGATGTTTCTTCTTCGATGCTTGAGGACCGTTCCGAACTCGA 436
DB 521 ATTTTGTTCGTCCTTTGAGTACGAGCAGCATTTGTTGGATGACCAAGACAAAGGCTC 580
OY 437 TTCACCGTGGGATGATGATCTTGTGATGTTGTGCGAGAGTGTTCGTAAGTTG 496
DB 581 TTCAATTCACATACATGAGATGGAAGCTTCTCTTGGATGATTTGAGAAATTAAT 640
OY 497 TTGATGACATCGTTCGGAATGTTGTATCTTCATGCTGGTATGATCCGTTGCACCTTG 556
DB 641 TTAATCCCCCACTTCCAATGATATATGTAATGCCATTGGTACTGATGCAATGG 700
OY 557 TTAAG 562
DB 701 ATAATG 706

US-09-740-035-3/C
Sequence 3, Application US/09740035
Patent No. 6344353
GENERAL INFORMATION:
APPLICANT: YE, Jane et al
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: CLO01058
CURRENT APPLICATION NUMBER: US/09/740,035
CURRENT FILING DATE: 2000-12-20
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 19736
TYPE: DNA
ORGANISM: Human
US-09-740-035-3

Query Match 1.6%; Score 36.8; DB 4; Length 19736;
Best Local Similarity 53.5%; Pred. No. 1.3;
Matches 77; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 1809 CCAAGAGAGAGATGTCGTCGAGACACGCCGCAACCAATGCTTTGACATTAACA 1868
DB 2368 CCTGGAGAGTGGAGTGTGTAGTACCTAAGACTGCACCACTGCACTCCAGCTGGGCAC 2309
OY 1869 ACAGTGCATCCAAACTAACCACCAAGACATAGATGCGACACTGGAGAAACCTAT 1928
DB 2308 AGAGGAAGACTCCGTCCTCAAAAAAAGAAAAAAGAAAGAGAGTGGAGTCCCTGT 2249
OY 1929 CTCTATTGAGGCGTTAAACCTT 1952
DB 2248 CCTCAGTGAAGAGCGCTTACAGCCT 2225

RESULT 7
US-09-010-928B-3
Sequence 3, Application US/09010928B
Patent No. 5994099
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V
APPLICANT: Hayashi, Cheryl Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
TITLE OF INVENTION: CODING THEREFOR
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 GATEHOUSE RD. SUITE 500E
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2824 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant

```

: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: FEATURE:
: NAME/KEY: -
: LOCATION: 1..2824
: OTHER INFORMATION: /note="Flagellin form DNA sequence
: OTHER INFORMATION: taken from 3' region. Stop codon begins at position 2722."
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..2724
: US-09-010-928B-3

Query Match 1.5%; Score 36; DB 2; Length 2824;
Best Local Similarity 52.7%; Pred. No. 0.63;
Matches 78; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

OY 1025 CGCAGTGGCTTGCTTGCCCTGAGGACACGTGTGATTACGCCAGTGTCTTACACAGGGCTGC 1084
Db 918 CGTAGACCTGGTGGTGCTGGTGAGACCTTACGGACCTGGGGGTTCTGACCTGAGGGTGC 977
OY 1085 AGATGACGATGGAGGATTTGGCCGTGATGAAGATTGTTCTTCATAGAGTCGGGCTGCC 1144
Db 978 AGCGGAGCTGGAGGACCTGGTGGAGCATACGGACCTGGTGATCATATGACCACTGGTGG 1037
OY 1145 CGCCGGAGAGGCTGTGAGGAGTGTACGA 1172
Db 1038 TTCGGAGGAGCCCGGTGTGCTGCGCGGA 1065

RESULT 8
: Sequence 8-630-4/c
: Sequence 4, Application US/09383630A
: Patent No. 6265632
: GENERAL INFORMATION:
: APPLICANT: Avner Yayon et al.
: TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH
: FACTOR RECEPTOR ASSOCIATED
: CHONDRODYSPLASIA
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Mark M. Friedman c/o Anthony Castlorina
: STREET: 2001 Jefferson Davis Highway, Suite 207
: CITY: Arlington
: STATE: Virginia
: COUNTRY: United States of America
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
: COMPUTER: Twinhead* Slimnote-890tx
: OPERATING SYSTEM: MS DOS version 3.11
: SOFTWARE: Word for Windows version 2.0 converted
: to an ASCII file
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/383,630A
: FILING DATE: 26-Aug-1999
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: <Unknown>
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Friedmam, Mark M.
: REGISTRATION NUMBER: 33,883
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 972-3-562553
: TELEFAX: 972-3-562554
: TELEX: <Unknown>
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8083
: TYPE: nucleic acid

```

```
; STRADEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
  
US-09-383-630-4
```

Query Match 1.5%; Score 36; DB 4; Length 8083;
Best Local Similarity 50.6%; Pred. No. 1.3;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

OY 742 TTAGGCGCAATGCAGACGTGGTCCGAATTCACATGCCAGATTCTCATCACCAA 801
DB 4977 TTTGTCCTAACCGCAGACTGTGGAAAAAAAGGGACAGATAGATATCATCATCAAG 4918
OY 802 GAAGAATGACAACCACCGTGCGCGAGCAATAATGCCCTGCGCTTTTGAATTCOA 861
DB 4917 GAAACCCGTGACTACTGCGCCCCGGGGAGATCTGCAGCCC CGGGGATCTTAAGATCCCT 4858
OY 862 GAAGAATTACAATTCGTCATCTACTCAATCCCAAGGCCACCGCCCGAG 913
DB 4857 AAGGGTTACTGTCCTCCACACTAACACACAGCAGCGGCTGCTCCTACCAAG 4806

RESULT 9
US-09-383-630-5/c
Sequence 5, Application US/09383630A
Patent No. 6265632
GENERAL INFORMATION:
APPLICANT: Avner Yaron et al.
TITLE OF INVENTION: ANIMAL MODEL FOR FIBROBLAST GROWTH FACTOR RECEPTOR ASSOCIATED CHONDRODYSPLASIA
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Mark M. Friedman c/o Anthony Castlorina
STREET: 2001 Jefferson Davis Highway, Suite 207
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
COMPUTER: Twinhead* Slimote-890TX
OPERATING SYSTEM: MS DOS version 6.2,
Windows version 3.11
SOFTWARE: Word for Windows version 2.0 converted
to an ASCII file
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/383,630A
FILING DATE: 26-Aug-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Friedmann, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 1402/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-5625553
TELEFAX: 972-3-5625554
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8083
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-383-630-5

Query Match 1.5%; Score 36; DB 4; Length 8083;
Best Local Similarity 50.6%; Pred. No. 1.3;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Db 1739 TACGGTAGAGSTGCTGGAGCTGGAGCTGCTGCAAGTGTGGAAGCTGAAGCCGCT 1798

QY 1149 GGAGAGGCTGAGAGATGTACGAAG 1174

Db 1799 GCAGGTGCAGAGACGAGAGCTGAGG 1824

RESULT 12
 US-08-425-069-1
 Sequence 1, Application US/08425069
 Patent No. 5728810
 GENERAL INFORMATION:
 APPLICANT: Lewis, Randolph V.
 APPLICANT: Xu, Ming
 APPLICANT: Himman, Michael B.
 TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
 TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
 TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
 NUMBER OF SEQUENCES: 69
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolasch & Birch
 STREET: 301 No. 5728810th Washington Street
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22046

Query Match	1.58;	Score 35.2;	DB 1;	Length 2338;
Best Local Similarity	49.08;	Pred. No. 1;		
Matches 94; Conservative	0;	Mismatches 98;	Indels 0;	Gaps 0

Accession	Sequence	Position
QY 1028	AGTGGCCCTGGCTTCCCTGAGAGCAGCGTATATGACGACGCTTTCACACAGGCGTGCAGA	1087
Db 1281	AGGTGGATTAGGTGAGCACAAGGGGAGGTGACAGCGCTGCACAGCATGGAGAGTGGCGG	13407
QY 1088	TGACGATGAGAGATTTGGCCGGTGGAATPAAGATTCCTTCATPAAGGTGGGGCTGTGCCGC	1147
Db 1341	ACAAGGAGGATATGAGAGGCCCTTGGAACAAGGTGCTGACGAGGGGAGCAAGAGTGCAGC	1400
QY 1148	CGGAGAAAGCGTGAGAGAGATATACGAAGGACCTTCATCTCATTTGGAAAAAGTTGGGATCG	1207
Db 1401	AGCAGCAGCTGGAGAGGTGCCGACACAAGAGAGATATGGAGGCTCTTTGGAAGGCCAAGAGTGTGG	1460
QY 1208	ACGGGTTTAAGA	1219
Db 1461	ACGAGGTGCACA	1472

RESULT 13
 US-08-317-844B-1
 Sequence 1, Application US/08317844B
 Patent No. 5989894
 GENERAL INFORMATION:
 APPLICANT: Lewis, Randolph V.
 APPLICANT: Xu, Ming
 APPLICANT: Himman, Michael B.
 TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
 TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
 TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
 NUMBER OF SEQUENCES: 62
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolasch & Birch
 STREET: 301 No. 5989894th Washington Street
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22046
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/317,844B
 FILING DATE: 04-OCT-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1447-105P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 241-1300
 TELEFAX: (703) 241-2848
 TELEX: 248345
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2338 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Nephila clavipes
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..2154
 OTHER INFORMATION: /product= "Nephila clavipes
 OTHER INFORMATION: dragline silk protein"
 PUBLICATION INFORMATION:
 AUTHORS: Xu, Ming
 AUTHORS: Lewis, Randolph V.
 TITLE: Structure of a protein superfiber: Spider

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 18:19:36 ; Search time 248 Seconds
(without alignments)
8318.946 Million cell updates/sec

Title: US-09-675-208-4_COPY_56_2407

Perfect score: 2352

Sequence: 1 ATGCCTCCTACTTTTAAAAA.....CGTTATCGACTTGTGT 2352

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCIT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCITUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	113.2	4.8	1036	10	US-09-770-445-180
2	86.2	3.7	1036	10	US-09-878-574-2803
3	75.4	3.2	296	10	US-09-294-0938-2999
4	45.8	1.9	377	10	US-09-878-574-1840
5	44.4	1.9	160	10	US-09-770-696-410
6	38	1.6	911	9	US-09-866-570A-19
7	38	1.6	911	9	US-09-866-570A-21
8	38	1.6	911	10	US-09-866-572A-19
9	38	1.6	911	10	US-09-866-572A-21
10	37.6	1.6	46050	10	US-09-820-003A-3
11	37.4	1.6	1664	9	US-10-091-504-1271
12	37.4	1.6	1664	9	US-10-091-504-1272
13	37.4	1.6	1664	10	US-09-764-869-1271
14	37.4	1.6	1664	10	US-09-764-869-1272
C 15	37.2	1.6	441	9	US-09-796-692-7801
16	37	1.6	5222	10	US-09-993-170-57
17	37	1.6	5696	10	US-09-993-170-60
18	37	1.6	5757	10	US-09-993-170-56
19	36.8	1.6	750	9	US-10-184-644-104

C 20	36.8	1.6	19736	12	US-10-014-502-3	Sequence 3, Appl1
21	36.2	1.5	485	10	US-09-864-761-11473	Sequence 11473, A
C 22	36	1.5	18554	10	US-09-811-825-3	Sequence 3, Appl1
23	35.4	1.5	2671	9	US-09-822-846-129	Sequence 129, Appl1
C 24	35.2	1.5	3159	9	US-10-071-766-85	Sequence 85, Appl1
25	35	1.5	653	9	US-10-184-644-402	Sequence 402, Appl
26	35	1.5	3919	9	US-10-025-384-1125	Sequence 1125, Ap
27	35	1.5	21761	9	US-10-092-154-1680	Sequence 1680, Ap
C 28	35	1.5	21761	10	US-09-764-847-1680	Sequence 1680, Ap
29	35	1.5	46718	9	US-10-274-873-3	Sequence 3, Appl1
C 30	35	1.5	46718	10	US-09-816-093-3	Sequence 3, Appl1
31	34.8	1.5	351	10	US-09-867-701-6593	Sequence 6593, Ap
C 32	34.8	1.5	671	9	US-10-184-644-346	Sequence 346, Appl
33	34.8	1.5	1326	9	US-09-866-570A-53	Sequence 53, Appl
C 34	34.8	1.5	1326	10	US-09-866-572A-53	Sequence 53, Appl
35	34.8	1.5	18596	9	US-09-954-531-124	Sequence 124, Appl
C 36	34.8	1.5	18596	9	US-09-954-531-124	Sequence 124, Appl
37	34.8	1.5	18596	10	US-09-880-107-1590	Sequence 1590, Ap
C 38	34.8	1.5	18596	10	US-09-967-768A-119	Sequence 119, Appl
39	34.6	1.5	378361	9	US-09-901-136-3	Sequence 3, Appl1
C 40	34.2	1.5	16181	9	US-10-092-154-1426	Sequence 1426, Ap
41	34.2	1.5	16181	10	US-09-764-847-1426	Sequence 1426, Ap
C 42	34	1.4	2139	10	US-09-815-242-6833	Sequence 6833, Ap
43	34	1.4	6259	10	US-09-070-927A-129	Sequence 129, Appl
C 44	34	1.4	10877	9	US-10-263-788-1	Sequence 1, Appl1
45	33.8	1.4	3087	10	US-09-815-242-7419	Sequence 7419, Ap

ALIGNMENTS

RESULT 1

US-09-770-445-180/C
Sequence 180, Application US/09770445
Patent No. US2002023281A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krickler, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2023US (PARA-012PRV)
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/178,472
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 180
LENGTH: 1036
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-445-180

Query Match 4.8%; Score 113.2; DB 10; Length 1036;
Best Local Similarity 62.9%; Pred. No. 6.2e-25;
Matches 175; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY	1570	TTCATCTGCTCTGTGAGGCATCTGGGAGCCCATATGTTAGATATCTGTGGCA	1629
Db	1036	TACCATGCTGCAGCGCTGAGTGGGTGGATGCGCATATGATGTACGTATAGCCAGGC	977
QY	1630	AAGCATACCTTGATCTCTTGAAAAAACAATGATCTTCTGATGATACGATCTTCGAAGT	1689
Db	976	AACCAACAACCTTGATCTATATTGAGGAAGCTGGTCTCTCCGATGATTCAGTTCTTCGGGCT	917
QY	1690	GAGTACTATGACATCCCGATCTGCGCATGTGGTGTGGAGAACCCCTTGATATAGGAGAA	1749
Db	916	AAGCTCCCGGTAGGCTTACCCTGTGACTGCTTATTTGCTGTATCCACGATAGAGATGGAATC	857
QY	1750	ACTATGCTTAAGATTTGGAACTCTCAACAAGTTCACTGAGATGTTGGTATCTCAATGC	1809
Db	856	AGCTTGGCTCAAGATCTGGAACATGATTAAGTTACTGATTTGGTGTATTCATATGT	797
QY	1810	CAAGGAGAGATGGTGTCTGTGAGAACGCCGCAACCA	1847
Db	796	CAAGTGCTGTTGTGTCTCAAGGAACGAAGAACCA	759

```

RESULT 2
US-09-878-574-2803
: Sequence 2803, Application US/09878574
: Patent No. US20020110548A1
: GENERAL INFORMATION:
: APPLICANT: Byrum, Joseph R.
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Thompson, Michael D.
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(15401)B
: CURRENT APPLICATION NUMBER: US/09/878,574
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 09/333,535
: PRIOR FILING DATE: 1999-06-14
: NUMBER OF SEQ ID NOS: 15775
: SEQ ID NO 2803
: LENGTH: 356
: TYPE: DNA
: ORGANISM: Glycine max
: OTHER INFORMATION: Clone ID: L1B3028-020-Q1-B1-B11
US-09-878-574-2803

```

	Query Match	3.7%	Score 86.12;	DB 10;	Length 356;
	Best Local Similarity	58.3%	Pred. No. 1e-16;		
Matches	151;	Conservative	0;	Mismatches 108;	Indels 0; Gaps 0
QY	1476	TCACATGGTTCATTGTGCCAACGACAGCTTGTGGATGGGAACTTCATCCACCCTGACTG	1535		
Db	91	TCACATTCACATCAGTGGCTTACAAATCTATTTTCCCTGTGATTAATATATCAGCCACACAG	150		
QY	1536	GGATATGTTCCATCCACCCACCTTGTGCCCTTCATCGCTGCCTCGAGGCATATC	1595		
Db	151	GGATATGTTTCTAATACCTACATCCGATGGCTGAATATCATGTGTGCAGCAGCTGTATG	210		
QY	1596	TGTGTGCCCCGATCTATGTTAGTATTTCTGTGGGAAAGCATTACTTTGATCTCTGAAAA	1655		
Db	211	GGGATGTCCAATTATATGTGAGTGAACAAGCCGACACCCATGTGACTTTGATCTTTTGAAGAA	270		
QY	1656	ACTAGTGCCTTCGTATGATCGATCGATCCTTCGAAAGTAGTACTATGCATCCGACCTGGCA	1715		
Db	271	GCTTGCACTACTGATGTGCTTATATTTAAGGCTTAACCTCCAGGACGACCAACAAAGGA	330		
QY	1716	TTGTTTGTGTAAGACCT	1734		
Db	331	TTGCTATTATCTATVCC	349		

RESULT 3
US-09-294-093B-2999
; Sequence 2999, Application US/09294093B

```

? Patent No. US20010051335A1
? GENERAL INFORMATION:
? APPLICANT: Lalgudi, Raghunath, V.
? APPLICANT: Ito, Laura, Y.
? APPLICANT: Sherman, Bradley, K.
? TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
? FILE REFERENCE: PI-0009 US
? CURRENT APPLICATION NUMBER: US/09/294,093B
? CURRENT FILING DATE: 1999-04-16
? PRIOR APPLICATION NUMBER: 60/082,567
? PRIOR FILING DATE: April 21, 1998
? NUMBER OF SEQ ID NOS: 6207
? SOFTWARE: PERL Program
? SEQ ID NO 2999
? LENGTH: 296
? TYPE: DNA
? ORGANISM: Zea mays
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: Incyte ID No. US20010051335A1 700347006H1
? NAME/KEY: unsure
? LOCATION: 190, 193-194, 206, 217, 232, 295
? OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-2999

Query Match 3.2% Score 75.4; DB 10; Length 296;
Best Local Similarity 66.1%; Pred. No. 2.4e-13;
Matches 109; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```

Oy	1521	CATCCACCCTACACGGGATATGTTCCAAATCCACCACCCCTTGTGGCGCTTCCATGATCGC	1580
Db	1	CATCGACGCACACGCGGACATGTTCCATAGTCTTACCCCATGGCGAATACCATCATGCTGC	60
Oy	1581	CTCTCGAGCACAATCTCTCTGGTGGCCCGATATATATTAAGTATGTCGTGGGAAACATATAT	1640
Db	61	GGCTCGAGCGTGTGTGCTGCTGTGCCATATATGTCCACGACAAAGCCTGGAGGCATGACTT	120
Oy	1641	TGATCTTCTGAAAAACTAGTCTCTCTGATGATGCATCTTCG	1685
Db	121	CAATCTGCTCAAGAGCTCTGCTGCTCCCGACGGATATGATCTCGCG	165

```

RESULT 4
US-09-878-574-1840
; Sequence 1840, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/733,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 1840
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-032-Q1-B1-E5
US-09-878-574-1840

```

	Query Match	1.9%	Score 45.8	DB 10	Length 377
	Best Local Similarity	55.3%	Pred. No.	0.00068	
	Matches	89	Conservative	0	Mismatches 72
				Incls	0
				Gaps	0
Oy	192	TGCTTCCTTTGTGGATTCGCACCGCTGGACACTGATAGCCCATGTTGTTTGCATTGG	251		
Db	217	TGGGCATTCATTCATGGGTTCAATCCATCACAAGGTACCGCACAGTCCTTCCCATTTTG	276		

FILE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091.504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1271
LENGTH: 1664
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1271

Query Match
Best Local Similarity 53.0%; Pred. No. 0.88;
Matches 80: Conservative 0; Mismatches 71: Indels 0; Gaps 0;

OY 1782 CACTGGAGTATTGGTGCATTTCACTGCCAAGAGAGAGAGTGTCTGTGAGACACGCCG 1841
DB 987 CACTCGGAGCGCTGAGCAGAGAGTGAACCCGGGAGCGAGGTTCAGTGCAGCGAGATC 1046
OY 1842 CAACCAATGCTTTTCAACAATCTCAAAAGAGTGCATCCAAACTAACCCAAAGACAT 1901
DB 1047 GCGCCAGTGCACCTTCAAGCTGGGCAACAAGAGGAACTCCTCAAAAAA 1106
OY 1902 AGAATGCGACAGTGGAGAAACCTATCTCT 1932
DB 1107 AGAAAAAGAAAAAGAAAGAAACTTTTCCT 1137

RESULT 12
US-10-091-504-1272
Sequence 1272, Application US/10091504
Publication No. US20030059908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091.504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1272
LENGTH: 1664
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1272

Query Match
Best Local Similarity 53.0%; Pred. No. 0.88;
Matches 80: Conservative 0; Mismatches 71: Indels 0; Gaps 0;

OY 1782 CACTGGAGTATTGGTGCATTTCACTGCCAAGAGAGAGAGTGTCTGTGAGACACGCCG 1841
DB 987 CACTCGGAGCGCTGAGCAGAGAGTGAACCCGGGAGCGAGGTTCAGTGCAGCGAGATC 1046
OY 1842 CAACCAATGCTTTTCAACAATCTCAAAAGAGTGCATCCAAACTAACCCAAAGACAT 1901
DB 1047 GCGCCAGTGCACCTTCAAGCTGGGCAACAAGAGGAACTCCTCAAAAAA 1106
OY 1902 AGAATGCGACAGTGGAGAAACCTATCTCT 1932
DB 1107 AGAAAAAGAAAAAGAAAGAAACTTTTCCT 1137

RESULT 13
US-09-764-869-1271
Sequence 1271, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764, 869
CURRENT FILING DATE: 2001-01-17
Prior Application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1271
LENGTH: 1664
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1271

Query Match
Best Local Similarity 53.0%; Pred. No. 0.88;
Matches 80: Conservative 0; Mismatches 71: Indels 0; Gaps 0;

OY 1782 CACTGGAGTATTGGTGCATTTCACTGCCAAGAGAGAGAGTGTCTGTGAGACACGCCG 1841
DB 987 CACTCGGAGCGCTGAGCAGAGAGTGAACCCGGGAGCGAGGTTCAGTGCAGCGAGATC 1046
OY 1842 CAACCAATGCTTTTCAACAATCTCAAAAGAGTGCATCCAAACTAACCCAAAGACAT 1901
DB 1047 GCGCCAGTGCACCTTCAAGCTGGGCAACAAGAGGAACTCCTCAAAAAA 1106
OY 1902 AGAATGCGACAGTGGAGAAACCTATCTCT 1932
DB 1107 AGAAAAAGAAAAAGAAAGAAACTTTTCCT 1137

RESULT 14
US-09-764-869-1272
Sequence 1272, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764, 869
CURRENT FILING DATE: 2001-01-17
Prior Application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1272
LENGTH: 1664
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1272

Query Match
Best Local Similarity 53.0%; Pred. No. 0.88;
Matches 80: Conservative 0; Mismatches 71: Indels 0; Gaps 0;

OY 1782 CACTGGAGTATTGGTGCATTTCACTGCCAAGAGAGAGAGTGTCTGTGAGACACGCCG 1841
DB 987 CACTCGGAGCGCTGAGCAGAGAGTGAACCCGGGAGCGAGGTTCAGTGCAGCGAGATC 1046
OY 1842 CAACCAATGCTTTTCAACAATCTCAAAAGAGTGCATCCAAACTAACCCAAAGACAT 1901
DB 1047 GCGCCAGTGCACCTTCAAGCTGGGCAACAAGAGGAACTCCTCAAAAAA 1106
OY 1902 AGAATGCGACAGTGGAGAAACCTATCTCT 1932
DB 1107 AGAAAAAGAAAAAGAAAGAAACTTTTCCT 1137

RESULT 15
US-09-796-692-7801/C
Sequence 7801, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane

```

? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
? TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
? FILE REFERENCE: 2077 001200
? CURRENT APPLICATION NUMBER: US/09/796,692
? CURRENT FILING DATE: 2001-03-01
? PRIOR APPLICATION NUMBER: 60/186,126
? PRIOR FILING DATE: 2000-03-01
? PRIOR APPLICATION NUMBER: 60/190,479
? PRIOR FILING DATE: 2000-03-17
? PRIOR APPLICATION NUMBER: 60/200,545
? PRIOR FILING DATE: 2000-04-27
? PRIOR APPLICATION NUMBER: 60/200,303
? PRIOR FILING DATE: 2000-04-28
? PRIOR APPLICATION NUMBER: 60/200,779
? PRIOR FILING DATE: 2000-04-28
? PRIOR APPLICATION NUMBER: 60/200,999
? PRIOR FILING DATE: 2000-05-01
? PRIOR APPLICATION NUMBER: 60/202,084
? PRIOR FILING DATE: 2000-05-04
? PRIOR APPLICATION NUMBER: 60/206,201
? PRIOR FILING DATE: 2000-05-22
? PRIOR APPLICATION NUMBER: 60/218,950
? PRIOR FILING DATE: 2000-07-14
? PRIOR APPLICATION NUMBER: 60/222,903
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: 60/223,416
? PRIOR FILING DATE: 2000-08-04
? PRIOR APPLICATION NUMBER: 60/223,378
? PRIOR FILING DATE: 2000-08-07
? NUMBER OF SEQ ID NOS: 9597
? SOFTWARE: FASTSEQ for Windows Version 3.0
? SEQ ID NO 7801
? LENGTH: 441
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-796-692-7801

```

Query Match	1.6%	Score 37.2	DB 9	Length 441
Best Local Similarity	57.9%	Pred. No. 0.4		
Matches	65	Conservative	0	Mismatches 48; Indels 0; Gaps 0
Oy	1809	CCAAAGGAGGAGATGTTGTCGTGAGACAGCCGCAACCAATGCTTTTCACATPACTCANA	1868	
Db	328	CCTGGAGGCGGAGGTTGTTGGTGAAGCCGAGATCGCACCAATTCATTTCCAGCTGGGGCAAC	269	
Oy	1869	ACGATGCAATCCAAAACCTAACCCAAAAACATAGATGACAGTGGAGAAA	1922	
Db	268	AAGACGGAACCTCCGCTTCAAAAAAGAAAGAAAGAAAAGAAAAGAAA	215	

Search completed: April 3, 2003, 21:26:43
Job time : 398 secs

GenCore version 5.1.4-P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:40 ; Search time 3.41463 Seconds
(without alignments)
120.634 Million cell updates/sec

Title: US-09-675-208-3

Perfect score: 80
Sequence: 1 YDODMWWVQVWMP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilltest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	14	US-08-846-234-3	Sequence 3, Appl1
2	80	100.0	784	US-08-846-234-5	Sequence 5, Appl1
3	38	47.5	404	US-09-302-769-21	Sequence 21, Appl1
4	38	47.5	527	US-09-659-166-2	Sequence 2, Appl1
5	38	47.5	528	US-08-802-466-2	Sequence 2, Appl1
6	38	47.5	528	US-09-350-484-2	Sequence 2, Appl1
7	38	47.5	1829	US-09-157-420-1	Sequence 1, Appl1
8	37	46.2	156	US-09-064-922-2	Sequence 2, Appl1
9	37	46.2	179	US-09-064-922-5	Sequence 5, Appl1
10	37	46.2	399	US-09-491-362-7	Sequence 7, Appl1
11	37	46.2	399	US-09-874-562-7	Sequence 7, Appl1
12	37	46.2	452	US-08-764-870-15	Sequence 16, Appl1
13	37	46.2	452	US-08-980-115-16	Sequence 16, Appl1
14	37	46.2	475	US-09-491-362-2	Sequence 2, Appl1
15	37	46.2	475	US-09-874-562-2	Sequence 2, Appl1
16	37	46.2	477	US-09-449-335-2	Sequence 6, Appl1
17	37	46.2	477	US-09-449-335-6	Sequence 10, Appl1
18	37	46.2	477	US-09-480-921B-10	Sequence 29, Appl1
19	37	46.2	477	US-09-480-921B-29	Sequence 11, Appl1
20	37	46.2	918	US-09-041-886-11	Sequence 265, Appl1
21	36	45.0	229	US-09-199-637A-265	Sequence 4, Appl1
22	36	45.0	400	US-09-184-001-4	Sequence 4, Appl1
23	36	45.0	404	US-09-184-001-2	Sequence 4, Appl1
24	36	45.0	3170	US-09-036-987A-4	Sequence 85, Appl1
25	35	43.8	18	US-09-370-700-4	Sequence 3298, Ap
26	35	43.8	18	US-08-484-192-85	
27	35	43.8	347	US-09-134-001C-3298	

28	35	43.8	473	1	US-08-597-236-13	Sequence 13, Appl1
29	35	43.8	473	1	US-08-746-682A-13	Sequence 13, Appl1
30	35	43.8	563	4	US-09-718-841-2	Sequence 2, Appl1
31	35	43.8	563	4	US-09-718-810-2	Sequence 2, Appl1
32	35	43.8	589	2	US-07-668-646-6	Sequence 6, Appl1
33	35	43.8	589	2	US-08-429-996-6	Sequence 6, Appl1
34	35	43.8	589	2	US-08-431-333-6	Sequence 6, Appl1
35	35	43.8	589	5	PCT-US91-02321-6	Sequence 6, Appl1
36	35	43.8	1385	1	US-07-876-280-2	Sequence 2, Appl1
37	35	43.8	1385	1	US-07-675-772-2	Sequence 2, Appl1
38	35	43.8	1385	1	US-08-063-170-2	Sequence 2, Appl1
39	35	43.8	1385	1	US-08-158-232-2	Sequence 2, Appl1
40	35	43.8	1385	1	US-08-304-626-2	Sequence 2, Appl1
41	35	43.8	1385	1	US-08-316-301A-2	Sequence 2, Appl1
42	35	43.8	1385	2	US-08-611-928-2	Sequence 2, Appl1
43	35	43.8	1385	3	US-09-173-891-2	Sequence 2, Appl1
44	35	43.8	1385	4	US-09-076-137-2	Sequence 2, Appl1
45	35	43.8	1385	5	PCT-US92-03624-2	Sequence 2, Appl1

ALIGNMENTS

```
RESULT 1
US-08-846-234-3
; Sequence 3, Application US/08846234
; Patent No. 6166292
;
GENERAL INFORMATION:
;
APPLICANT: OSUMI Chieko
;
APPLICANT: NOZAKI Jinshi
;
APPLICANT: KIDA Takao
;
TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
;
TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
;
NUMBER OF SEQUENCES: 22
;
CORRESPONDENCE ADDRESS:
;
ADDRESS: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT, P. C.
;
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
;
CITY: ARLINGTON
;
STATE: VIRGINIA
;
COUNTRY: USA
;
ZIP: 22202
;
COMPUTER READABLE FORM:
;
MEDIUM TYPE: Floppy disk
;
COMPUTER: IBM PC compatible
;
OPERATING SYSTEM: PC-DOS/MS-DOS
;
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;
CURRENT APPLICATION DATA:
;
APPLICATION NUMBER: US/08/846,234
;
FILING DATE:
;
ATTORNEY/AGENT INFORMATION:
;
NAME: NORMAN F. OBLON
;
REGISTRATION NUMBER: 24 618
;
TELECOMMUNICATION INFORMATION:
;
TELEPHONE: (703)-413-3000
;
TELEFAX: (703)-413-2220
;
INFORMATION FOR SEQ ID NO: 3:
;
SEQUENCE CHARACTERISTICS:
;
LENGTH: 14 amino acids
;
TYPE: amino acid
;
TOPOLOGY: linear
;
MOLECULE TYPE: peptide
;
FRAGMENT TYPE: Internal
;
US-08-846-234-3
;
Query Match 100.0%; Score 80; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 YDODMWWVQVWMP 14
Db 1 YDODMWWVQVWMP 14
```

```
RESULT 2
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-846-234-5

Query Match          100.0%; Score 80; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 5.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 YDDQDMVVVQVWP 14
        ||| ||| ||| |||
Db       756 YDDQDMVVVQVWP 769

RESULT 3
US-09-302-769-21
; Sequence 21, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; FILE REFERENCE: 109762
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; CURRENT APPLICATION NUMBER: US/09/302,769
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 404
```

```
; TYPE: PRT
; ORGANISM: Human
US-09-302-769-21

Query Match          47.5%; Score 38; DB 4; Length 404;
Best Local Similarity 50.0%; Pred. No. 58;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY      1 YDDQDMVVVQVWP 14
        : | || |||
Db       43 WSGCHGVKIVWP 56

RESULT 4
US-09-659-166-2
; Sequence 2, Application US/09659166
; Patent No. 6355465
; GENERAL INFORMATION:
; APPLICANT: GREASY, CAREYTHA LEE
; APPLICANT: TESTA, TANIA TAMSON
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30218
; CURRENT APPLICATION NUMBER: US/09/659,166
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: UK 9921505.5
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 527
; TYPE: PRT
; ORGANISM: rattus
; US-09-659-166-2

Query Match          47.5%; Score 38; DB 4; Length 527;
Best Local Similarity 66.7%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 YDDQDMVVVQVWP 12
        || || |||
Db       129 YDDQGSYQVWP 140

RESULT 5
US-08-802-466-2
; Sequence 2, Application US/08802466
; Patent No. 5972606
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/08/802,466
; FILING DATE: 19 February 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
```

REFERENCE/DOCKET NUMBER: GH50002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-802-466-2

Query Match 47.5%; Score 38; DB 2; Length 528;
Best Local Similarity 66.7%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 YDDQDMVVQVP 12
1111111111
Db 130 YDDQDGSYVQVP 141

RESULT 6
US-09-350-484-2
Sequence 2, Application US/09350484
Patent No. 6159716
GENERAL INFORMATION:
APPLICANT: Creasy, et al.
TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/350,484
FILING DATE: 09-Jul-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/802,466
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: William T. Han
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: GH50002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-350-484-2

Query Match 47.5%; Score 38; DB 4; Length 528;
Best Local Similarity 66.7%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 YDDQDMVVQVP 12

Db 130 YDDQDGSYVQVP 141
1111111111

RESULT 7
US-09-157-420-1
Sequence 1, Application US/09157420
Patent No. 6180760
GENERAL INFORMATION:
APPLICANT: TAKAI, Yoshimi
APPLICANT: NAKANISHI, Hiroyuki
APPLICANT: MANDAI, Kenji
APPLICANT: WADA, Manabu
APPLICANT: OBAISHI, Hiroshi
TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-APADIN"
FILE REFERENCE: 98-1042/LC(WMC)/653
CURRENT APPLICATION NUMBER: US/09/157,420
CURRENT FILING DATE: 1998-09-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1829
TYPE: PRT
ORGANISM: rat
US-09-157-420-1

Query Match 47.5%; Score 38; DB 4; Length 1829;
Best Local Similarity 58.3%; Pred. No. 3.4e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 QDMVVQVQVMP 14
1111111111
Db 1246 QDMAPVQVQVMP 1257

RESULT 8
US-09-064-922-2
Sequence 2, Application US/09064922
Patent No. 622095
GENERAL INFORMATION:
APPLICANT: Callis, Judy
APPLICANT: Worley, Cathy K.
TITLE OF INVENTION: Sequences from Auxin-Induced Gene
TITLE OF INVENTION: Products Targeting Fusion Proteins for Degradation
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,922
FILING DATE: 22-Apr-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Laurence J.
REGISTRATION NUMBER: 35,551
REFERENCE/DOCKET NUMBER: 023070-085400US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
TOPOLOGY: linear


```

; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,870
; FILING DATE: 13-DEC-1996
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,540
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,543
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,606
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Jackie N
; REGISTRATION NUMBER: 35,966
; REFERENCE/DOCKET NUMBER: UCAL-246/010US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)843-5000
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-764-870-16

Query Match      46.2%; Score 37; DB 4; Length 452;
Best Local Similarity 60.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 DOMVVVQVPM 13
    1111111111
Db 265 DOMAVIQYSM 274

RESULT 13
; US-08-980-115-16
; Sequence 16, Application US/08980115
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S.
; APPLICANT: Baxter, John D.
; APPLICANT: Rietterick, Robert J.
; APPLICANT: Wagner, Richard L.
; APPLICANT: Kushner, Peter J.
; APPLICANT: Apriettl, James W.
; APPLICANT: West, Brian L.
; APPLICANT: Shlau, Andrew K.
; TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
; FILE REFERENCE: UCAL-246/020US
; CURRENT APPLICATION NUMBER: US/08/980,115
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: 08/764,870
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 60/008,606
; EARLIER FILING DATE: 1995-12-14
; EARLIER APPLICATION NUMBER: 60/008,543
; EARLIER FILING DATE: 1995-12-13
; EARLIER APPLICATION NUMBER: 60/008,540
; EARLIER FILING DATE: 1995-12-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 16
; LENGTH: 452
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (184)..(437)
; OTHER INFORMATION: minimal ligand binding domain
; US-08-980-115-16

Query Match      46.2%; Score 37; DB 4; Length 452;
Best Local Similarity 60.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 DOMVVVQVPM 13
    1111111111
Db 265 DOMAVIQYSM 274

RESULT 14
; US-09-491-362-2
; Sequence 2, Application US/09491362
; Patent No. 6281017
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
; FILE REFERENCE: WSUR14977
; CURRENT APPLICATION NUMBER: US/09/491,362
; CURRENT FILING DATE: 2000-01-26
; EARLIER APPLICATION NUMBER: 60/118,349
; EARLIER FILING DATE: 1999-02-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Mentha piperita
; US-09-491-362-2

Query Match      46.2%; Score 37; DB 4; Length 475;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 QDOMVVVQVPM 14
    1111111111
Db 331 QDSSVLAQLGMP 342

RESULT 15
; US-09-874-562-2
; Sequence 2, Application US/09874562
; Patent No. 6420159
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; TITLE OF INVENTION: 1-DEOXY-D-XYLULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
; FILE REFERENCE: WSUR17549
; CURRENT APPLICATION NUMBER: US/09/874,562
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/491,362
; PRIOR FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: 60/118,349
; PRIOR FILING DATE: 1999-02-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Mentha piperita
; US-09-874-562-2

Query Match      46.2%; Score 37; DB 4; Length 475;
```

Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 ODOMVVOYVWP 14
11 1: 1: 11
Db 331 QDSSVLAQLGWP 342

Search completed: March 27, 2003, 10:07:35
Job time : 4.41463 secs

CC Raffinose synthase forms raffinose from sucrose and galactinol, has
 CC an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees
 CC C, has a molecular weight of 75 to 95 kDa by gel filtration or 90
 CC to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is
 CC inhibited by Iodacetamide, N-ethylmaleimide and myoinositol.

XX
 SQ Sequence 14 AA:

Query Match 100.0%; Score 80; DB 19; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVQVWP 14
 DB 1 YDQDMVVQVWP 14

RESULT 2

AA17424
 ID AAY17424 standard; peptide; 14 AA.

XX
 AC AAY17424;

XX
 DT 29-JUL-1999 (first entry)

XX
 DE Cucumber raffinose synthase peptide SEQ ID NO:3.

XX
 KW Raffinose synthase; sucrose; galactinol.

XX
 OS Cucumis sativus.

XX
 PN JP1123080-A.

XX
 PD 11-MAY-1999.

XX
 PF 24-OCT-1997; 97JP-0292969.

XX
 PR 24-OCT-1997; 97JP-0292969.

XX
 PA (AJIN) AJINOMOTO KK.

XX
 DR WPI: 1999-340516/29.

XX
 PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

XX
 PS Example 2; Page 22; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents a raffinose
 CC synthase peptide from cucumber.

XX
 SQ Sequence 14 AA:

Query Match 100.0%; Score 80; DB 20; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVQVWP 14
 DB 1 YDQDMVVQVWP 14

RESULT 3

AAW53570
 ID AAW53570 standard; Protein; 784 AA.

XX
 AC AAW53570;

XX
 DT 06-JUL-1998 (first entry)

XX
 DE Cucumber raffinose synthase.

XX
 KW Cucumber; raffinose synthase; sucrose; galactinol.

XX
 OS Cucumis sativus.

XX
 PN JP10084973-A.

XX
 PD 07-APR-1998.

XX
 PF 28-APR-1997; 97JP-0111124.

XX
 PR 26-JUL-1996; 96JP-0198079.

XX
 PR 26-APR-1996; 96JP-0107682.

XX
 PA (AJIN) AJINOMOTO KK.

XX
 DR WPI: 1998-264858/24.

XX
 DR N-PSDB; AAW22250.

XX
 PT Raffinose synthase gene - useful for preparation of raffinose in
 PT transformed plant

XX
 PS Claim 3; Pages 17-20; 26pp; Japanese.

CC The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC Iodacetamide, N-ethylmaleimide and myoinositol.

XX
 SQ Sequence 784 AA:

Query Match 100.0%; Score 80; DB 19; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVQVWP 14
 DB 756 YDQDMVVQVWP 769

RESULT 4

AA17417
 ID AAY17417 standard; Protein; 784 AA.

XX
 AC AAY17417;

XX
 DT 29-JUL-1999 (first entry)

XX
 DE Cucumber raffinose synthase.

XX
 KW Raffinose synthase; sucrose; galactinol.

XX
 OS Cucumis sativus.

XX
 PN JP1123080-A.

XX
 PD 11-MAY-1999.

XX
 PF 24-OCT-1997; 97JP-0292969.

XX
 PR 24-OCT-1997; 97JP-0292969.

XX
 PA (AJIN) AJINOMOTO KK.

XX
 DR WPI: 1999-340516/29.

XX
 DR N-PSDB; AAX61238.

XX
 PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

XX Claim 2: Page 25-27; 37pp; Japanese.
PS
XX
CC The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from cucumber.
XX
SQ Sequence 784 AA;

Query Match 100.0%; Score 80; DB 20; Length 784;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVVQVWMP 14
DB 756 YDQDMVVVQVWMP 769

RESULT 5
AAB98659
ID AAB98659 standard; protein; 780 AA.
XX
AC AAB98659;
XX

DT 17-AUG-2001 (first entry)
XX
DE Soybean protein: SEQ ID 1.
XX
KW Mutant; mutein; raffinose synthase; raffinose oligosaccharide reduction;
XX plant; soybean.
XX
OS Glycine max.
XX
PN JP2001078783-A.
XX
PD 27-MAR-2001.
XX
PF 03-JUL-2000; 2000JP-0200571.
XX
PR 09-JUL-1999; 99JP-0196036.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI: 2001-313373/33.
XX
DR N-PSDB: AAH27438.
XX

PT Novel mutant protein of raffinose synthase is useful for reducing the
XX raffinose oligosaccharide content in a plant body -
XX
PS Disclosure: Page 18-20; 30pp; Japanese.
XX

CC The present invention relates to a mutant protein of raffinose synthase
CC in which at least one aromatic amino acid present at the position of
CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
CC mutant protein can be used for reducing the raffinose oligosaccharide
CC content in a plant body. The present protein from soybean, was used in
CC the present invention.
XX
SQ Sequence 780 AA;

Query Match 62.5%; Score 50; DB 22; Length 780;
Best Local Similarity 66.7%; Pred. No. 2.9;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ODOMVVVQVWMP 14
DB 757 EDKMLRVQVWMP 768

RESULT 6

AAW57887
ID AAW57887 standard; protein; 781 AA.
XX

AC AAW57887;
XX
DT 23-SEP-1998 (first entry)
XX
DE Soybean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
XX gastrointestinal flora; soybean.
XX
OS Glycine max.
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.
XX
PF 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Wantanabe E;
XX
DR WPI: 1998-324670/29.
XX
DR N-PSDB: AAV40801.
XX

PT New nucleic acid molecule encoding plant raffinose synthetase -
XX capable of producing raffinose, used as food additives with
XX beneficial effects on gastrointestinal flora
XX
PS Claim 1: Page 31-34; 44pp; English.
XX

CC This sequence represents the soybean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX

SQ Sequence 781 AA;

Query Match 62.5%; Score 50; DB 19; Length 781;
Best Local Similarity 66.7%; Pred. No. 2.9;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ODOMVVVQVWMP 14
DB 758 EDKMLRVQVWMP 769

RESULT 7
AAV30143
ID AAV30143 standard; protein; 781 AA.
XX

AC AAV30143;
XX
DT 26-OCT-1999 (first entry)
XX
DE Amino acid sequence of a raffinose synthase protein.
XX
KW Raffinose synthase; plant; sucrose; raffinose.
XX
OS Glycine max.
XX
PN JP11215984-A.
XX
PD 10-AUG-1999.

XX 12-DEC-1997: 97JP-0342899.
XX
XX 28-NOV-1997: 97JP-0329006.
PR 18-DEC-1996: 96JP-0338673.
XX
XX (SUMO) SUMITOMO CHEM CO LTD.
XX
XX WPI: 1999-51112/43.
DR N-PSDB: AAC10002.
XX
XX New raffinose synthase gene - is prepared from a plant material
PT
XX
XX Claim 8: Page 25-27; 40pp: Japanese.
XX
XX The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6)-D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
SQ Sequence 781 AA;

Query Match 62.5%; Score 50; DB 20; Length 781;
Best Local Similarity 66.7%; Pred. No. 2.9;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ODOMVVVQVWP 14
: : : : :
Db 758 EDKMLRVQVWP 769

RESULT 8
AAB49400
ID AAB49400 standard; Protein: 781 AA.
XX

AC AAB49400;

XX 07-MAR-2001 (first entry)

XX Soybean raffinose synthase.

XX Plant promoter; transgenic plant; desired property.

XX Glycine max.

XX EP1048733-A2.

XX 02-NOV-2000.

XX 27-APR-2000; 2000EP-0108962.

XX 30-APR-1999; 99JP-0124527.

PR 01-SEP-1999; 99JP-0247211.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX Ishige F, Watanabe E, Oeda K;

DR WPI: 2001-104537/12.

DR N-PSDB: AAC89523.

XX New soybean plant promoters useful for generating transgenic plants

PT with desired properties -

XX Example 6; Page 24-27; 36pp: English.

XX The present invention provides novel plant promoters which can be used in

CC the production of transgenic plants which express genes with desired

XX properties.
SQ Sequence 781 AA;

Query Match 62.5%; Score 50; DB 22; Length 781;

Best Local Similarity 66.7%; Pred. No. 2.9;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 3 ODOMVVVQVWP 14
: : : : :
Db 758 EDKMLRVQVWP 769

RESULT 9
ABB93664
ID ABB93664 standard; Protein: 783 AA.
XX

AC ABB93664;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 2875.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidler M;

XX WPI: 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -

XX Claim 5; SEQ ID NO 2875; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins

XX (ABB90790-ABB94016) for herbicidally active compounds, comprising

XX aligning and comparing nucleic acid or amino acid sequences from plant

XX with nucleic acid or amino acid sequences from non-plant organisms using

XX suitable search parameters, where plant sequences having an E-value

XX greater by a factor of 3 than the E-value of most similar non-plant

XX sequences are selected. The polypeptides or nucleic acids encoding them

XX are useful for identifying modulators. The identified modulators are

XX useful as herbicides.

SQ Sequence 783 AA;

Db 760 EDSMVVQVWP 770

OY 3 ODOMVVVQVWP 13
: : : : :
Db 760 EDSMVVQVWP 770

RESULT 10
AAU67218
ID AAU67218 standard; Protein: 98 AA.

XX AAU67218;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #28114.
DE

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
OS
XX WO200181581-A2.
PN
XX 01-NOV-2001.
PD
XX 20-APR-2001; 2001WO-US12865.
PF
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
PA
XX Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
PI
XX WPI: 2001-616774/71.
DR N-PSDB: AAS59786.
DR
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1: SEQ ID NO 28413; 1069pp; English.
PS
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 98 AA:
Query Match 57.5%; Score 46; DB 22; Length 98;
Best Local Similarity 77.8%; Pred. NO. 1.5;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 6 MVVVOVPWP 14
I I I I I I I I
Db 9 MTVVOIPWP 17

RESULT 11
AAU49825
ID AAU49825 standard; Protein: 148 AA.
XX
XX AAU49825;
AC
XX 27-FEB-2002 (first entry)
DT
XX
XX Propionibacterium acnes immunogenic protein #10721.
DE

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
OS
XX WO200181581-A2.
PN
XX 01-NOV-2001.
PD
XX 20-APR-2001; 2001WO-US12865.
PF
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
PA
XX Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
PI
XX WPI: 2001-616774/71.
DR N-PSDB: AAS59545.
DR
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1: SEQ ID NO 11020; 1069pp; English.
PS
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 148 AA:
Query Match 57.5%; Score 46; DB 22; Length 148;
Best Local Similarity 77.8%; Pred. NO. 2.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 6 MVVVOVPWP 14
I I I I I I I I
Db 9 MTVVOIPWP 17

RESULT 12
AAG53157
ID AAG53157 standard; Protein: 484 AA.
XX
XX AAG53157;
AC
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 67650.
DE

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 17-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.

PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 02-JUL-1999; 990S-0142154.
PR 06-JUL-1999; 990S-0142055.
PR 08-JUL-1999; 990S-0142390.
PR 09-JUL-1999; 990S-0142803.
PR 12-JUL-1999; 990S-0142920.
PR 13-JUL-1999; 990S-0142977.
PR 14-JUL-1999; 990S-0143542.
PR 15-JUL-1999; 990S-0143624.
PR 16-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144684.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0148368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 56.2%; Score 45; DB 21; Length 484;
Best Local Similarity 53.8%; Pred. No. 13;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 DDDQMVVVOVWPP 14
1:1:1:1 111

DB 457 DEDDEMLVGDDPWP 469

RESULT 13

MAC53156
ID AAG53156 standard; Protein; 563 AA.

AC AAG53156;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 67649.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.

```

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

```

```

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          56.2%; Score 45; DB 21; Length 563;
Best Local Similarity 53.8%; Pred No. 15;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      2 DDDQWVVQVPP 14
       1:1:1:1:1
Db      536 DDEDEMLVGDDPP 548

RESULT 14
AAV32075
ID      AAV32075 standard; Protein; 572 AA.
XX
AC      AAV32075;
XX
DT      17-JAN-2000 (first entry)
XX
DE      Rapeseed raffinose synthase.
XX
KW      Raffinose synthase; rapeseed; transgenic plant.
XX
OS      Brassica napus.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 129 /note= "encoded by GCY"
FT      Misc-difference 132 /note= "encoded by GSW"
FT      Misc-difference 143 /note= "encoded by CCS"
FT      Misc-difference 144 /note= "encoded by CCS"
FT      Misc-difference 144 /note= "encoded by TCR"

```

FT Misc-difference 148
/note= "encoded by CCR"
FT
XX EP953643-A2.
PN 03-NOV-1999.
XX
PD 27-APR-1999; 99EP-0107430.
XX
PF 30-APR-1998; 98JP-0120550.
PR 30-APR-1998; 98JP-0120551.
PR 04-DEC-1998; 98JP-0345590.
PR 10-DEC-1998; 98JP-0351246.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Watanabe E, Oeda K;
DR WPI; 1999-593144/51.
DR N-PSDB; AA220210.
XX
XX New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
PT
PS Claim 27; Page 36-38; 55pp; English.
XX
XX This sequence represents rapessed raffinose synthase, a protein
CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC hydroxy group attached to the carbon atom at the 6-position of the
CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
CC (see AA220210) encoding the enzyme was isolated from rapessed cv.
CC Westar leaf cDNA by PCR. Probes or primers generated from plant
CC raffinose synthase genes (see AA220207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
XX
SQ Sequence 572 AA:

Query Match 56.2%; Score 45; DB 20; Length 572;
Best Local Similarity 63.6%; Pred. No. 16;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 ODOMVVVVVPPW 13
::|:|||||
Db 549 EESWVWVQVPM 559

RESULT 15
AAG53155
ID AAG53155 standard; Protein; 586 AA.
XX
AC AAG53155;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 67648.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.

XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145386.
PR 02-AUG-1999; 99US-0145388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 56.2%; Score 45; DB 21; Length 586;
Best Local Similarity 53.8%; Pred. No. 16;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 DDDQVVVGVMP 14
DB 559 DEDDMVGDPP 571

Search completed: March 27, 2003, 10:03:44
Job time : 10.9024 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 4.0122 Seconds
(without alignments)
335.448 Million cell updates/sec

Title: US-09-675-208-3
Perfect score: 80
Sequence: 1 YDQDMVYVQVWPP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	55.0	164	2 S58501	auxin-induced prot
2	44	55.0	225	2 H71407	auxin-induced prote
3	44	55.0	229	2 H86173	hypothetical prote
4	44	55.0	234	2 C85159	IAA7 like protein
5	44	55.0	619	2 C86467	hypothetical prote
6	44	55.0	719	2 T47727	hypothetical prote
7	43	53.8	418	2 G97360	intracellular PHB
8	43	53.8	418	2 AH2578	intracellular PHB
9	43	53.8	423	2 T52251	5-adenylylsulfate
10	43	53.8	863	1 S51789	VLDL receptor prec
11	42	52.5	500	2 D86274	hypothetical prote
12	42	52.5	529	2 T42584	tegument protein 4
13	42	52.5	600	2 C86468	probable auxin res
14	42	52.5	767	2 C86476	protein F1504.37 l
15	41	51.2	530	1 W2BED4	gene 40 protein -
16	40	50.0	189	2 S58493	auxin-induced prot
17	40	50.0	189	2 E95659	probable IAA6 prot
18	40	50.0	346	2 AG2094	hypothetical prote
19	40	50.0	448	2 F95036	pts system, IIC co
20	40	50.0	448	2 B97907	phosphotransferase
21	40	50.0	653	2 S75763	acetate-CoA ligase
22	40	50.0	1159	2 T43461	probable phosphodi
23	39	48.8	206	2 S22400	insecticynin A -
24	39	48.8	281	2 A65219	phn protein - Esc
25	39	48.8	281	2 A91264	phosphonate metabo
26	39	48.8	281	2 F86104	phosphonate metabo
27	39	48.8	305	2 T08963	phytochrome-associ
28	39	48.8	384	2 G82670	general secretory
29	39	48.8	848	2 JG0194	androgen receptor

30	38	47.5	162	2 C82765	conserved hypothet
31	38	47.5	190	2 C64950	yeam protein - Esc
32	38	47.5	190	2 A09952	hypothetical prote
33	38	47.5	190	2 B85800	hypothetical prote
34	38	47.5	305	2 A25010	pectinesterase (EC
35	38	47.5	390	2 A70656	hypothetical prote
36	38	47.5	432	2 S22527	glutamate-amonia
37	38	47.5	473	2 E84471	probable beta-1,3-
38	38	47.5	524	2 D82220	conserved hypothet
39	38	47.5	569	2 C86934	probable membrane
40	38	47.5	624	2 B96536	hypothetical prote
41	38	47.5	722	2 T30995	hypothetical prote
42	38	47.5	1206	2 S72620	probable reverse t
43	38	47.5	1663	2 T42092	s-afadin - rat
44	38	47.5	1687	2 T30244	phosphodiesterase
45	38	47.5	1706	2 T30175	exoribonuclease, v

ALIGNMENTS

RESULT 1
S58501
auxin-induced protein IAA14 - Arabidopsis thaliana (fragment)
N:Alternate names: indoleacetic acid-inducible protein IAA14
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 20-Aug-1999
C:Accession: S58501; S71408
R:Abel, S.; Nguyen, M.D.; Theologis, A.
J. Mol. Biol. 251, 533-549, 1995
A>Title: The ps-IAA4/5-like family of early auxin-inducible mRNAs in Arabidopsis thal
A:Reference number: S58491; MUID:9538793; PMID:7658471
A:Accession: S58501
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-164 <ABE>
R:Theologis, A.
submitted to the EMBL Data Library, December 1994
A:Reference number: S71406
A:Accession: S71408
A:Molecule type: DNA
A:Residues: 1-109, 'S', 111-164 <THE>
A:Cross-references: EMBL:U18416; NID:g972930; PIDN:AAC49055.1; PID:g972931
C:Genetics:
A:Gene: IAA14
C:Superfamily: auxin-induced protein aux28

Query Match 55.0%; Score 44; DB 2; Length 164;
Best Local Similarity 61.5%; Pred. No. 2;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDMVYVQVWPP 14
DB 120 DQDMVYVQVWPP 132

RESULT 2
H71407
auxin-induced protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C>Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C:Accession: H71407
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Ertlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Putdomen
ehoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalwatzis, N.
A>Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:9812113; PMID:9461215
A:Accession: H71407

G97360
Intracellular PHB depolymerase (AB017612) [Imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: G97360
R:Goodner, B.; Mollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Matkeiz, B.;
Science 294, 2223-2326, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: G97360
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-418 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK85840.1; PID:G15154875; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_24
A:Map position: circular chromosome
C:Superfamily: Rickettsia prowazekii hypothetical protein RP681

Query Match 53.8%; Score 43; DB 2; Length 418;
Best Local Similarity 87.5%; Pred. No. 8.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 7 VVQVQVWP 14
11:111111

Db 242 VVQVQVWP 249

RESULT 8
AH2578
Intracellular PHB depolymerase [Imported] - Agrobacterium tumefaciens (strain C58, Dupont
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AH2578
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, L.
erage, G.; Gillet, M.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Kemp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH2578
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-418 <KUR>
A:Cross-references: GB:AE006888; PIDN:AAI41046.1; PID:G17738332; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: dephA
A:Map position: circular chromosome
C:Superfamily: Rickettsia prowazekii hypothetical protein RP681

Query Match 53.8%; Score 43; DB 2; Length 418;
Best Local Similarity 87.5%; Pred. No. 8.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 7 VVQVQVWP 14
11:111111

Db 242 VVQVQVWP 249

RESULT 9
T52251
5'-adenylylsulfate reductase (EC 1.8.99.-) [validated] - green alga (Enteromorpha intest
C:Species: Enteromorpha intestinaalis (hollow green seaweed)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T52251
R:Guo, Y.; Schofield, O.M.; Leustek, T.
Plant Physiol. 123, 1087-1096, 2000
A:Title: Characterization of sulfate assimilation in marine algae focusing on the enzyme
A:Reference number: Z26003; MUID:20349720; PMID:10889258

A:Accession: T52251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-423 <GAO>
A:Cross-references: EMBL:AF069951; PIDN:AAK26855.1
A:Function:
A:Description: catalyzes reduced glutathione-dependent reduction of adenylylsulfate t
C:Keywords: oxidoreductase

Query Match 53.8%; Score 43; DB 2; Length 423;
Best Local Similarity 58.3%; Pred. No. 8.9;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 DQDQVVVQVWP 13
1:111111

Db 330 DQDQVVVQVWP 341

RESULT 10
S51789
VLDL receptor precursor - chicken
N:Alternate names: very low density lipoprotein receptor; vitellogenin receptor
C:Species: Gallus gallus (chicken)
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 16-Jul-1999
C:Accession: S51789
R:Bujo, H.; Hermann, M.; Kaderli, M.O.; Jacobsen, L.; Sugawara, S.; Nimpf, J.; Yamano
EMBO J. 13, 5165-5175, 1994
A:Title: Chicken oocyte growth is mediated by an eight ligand binding repeat member o
A:Reference number: S51789; MUID:95045409; PMID:7957081
A:Accession: S51789
A:Molecule type: mRNA
A:Residues: 1-863 <BU>
A:Cross-references: EMBL:X80207; NID:9609265; PIDN:CAA56505.1; PID:9609266
C:Comment: This receptor mediates uptake of very low density lipoprotein and vitellog
C:Comment: There is some evidence for the existence of a longer splice form contain
C:Keywords: duplication; fatty acid metabolism; glycoprotein; receptor; transmembrane
F:1-44/Domain: signal sequence #status predicted <SIG>
F:45-863/Product: VLDL receptor #status predicted <EXT>
F:45-787/Domain: extracellular #status predicted <EXT>
F:51-85/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:90-126/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:131-167/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:172-206/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:211-247/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:257-291/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:296-330/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:336-373/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F:378-412/Domain: EGF homology <EG1>
F:418-452/Domain: EGF homology <EG2>
F:459-498/Domain: LDL receptor WTD-containing repeat homology <YW1>
F:499-544/Domain: LDL receptor WTD-containing repeat homology <YW2>
F:545-587/Domain: LDL receptor WTD-containing repeat homology <YW3>
F:588-631/Domain: LDL receptor WTD-containing repeat homology <YW4>
F:632-674/Domain: LDL receptor WTD-containing repeat homology <YW5>
F:675-717/Domain: LDL receptor WTD-containing repeat homology <YW6>
F:726-769/Domain: EGF homology <EG3>
F:788-809/Domain: transmembrane #status predicted <TM>
F:824-828/Region: coated-pit mediated internalization signal
F:159-773/Binding site: carbohydrate (4sn) (covalent) #status predicted
F:378-389,385-398,400-412,418-428,437,439-452,726-735-734,736-769/Disulfide b

Query Match 53.8%; Score 43; DB 1; Length 863;
Best Local Similarity 35.7%; Pred. No. 20;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 YDQDQVVVQVWP 14
1:111111

Db 609 YDQDQVVVQVWP 622

RESULT 11
D86274

hypothetical protein F7A19.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence-Revision 02-Mar-2001 #text-change 31-Dec-2001
C:Accession: D86274
R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hutzlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Liu, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzall,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86274
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-500 <STO>
A:Cross-references: GB:AE005172; NID:95080784; PIDN:AAD39294.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match	52.5%	Score 42:	DB 2:	Length 500:
Best Local Similarity	50.0%:	Pred. NO. 16:		
Matches 6, Conservative	4:	Mismatches 2:	Indels 0:	Gaps 0:

Qy	2	DQDOMVVVQVPW	13
		: : : : :	
Db	225	ESDQILIRQVPW	236

RESULT 12
T42584
tegument protein 40 - equine herpesvirus 4 (strain NS80567)
C:Species: equine herpesvirus 4

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42584
R:Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
J. gen. Virol. 79, 1197-1203, 1998
A:Title: The DNA sequence of equine herpesvirus-4.
A:Reference number: 222173; MUID:98264497; PMID:9603335
A:Accession: T42584
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-529 <TEL>
A:Cross-references: EMBL:AF030027; NID:g2605950; PIDN:AAC59558.1; PID:g2605986
A:Experimental source: strain NS80567
C:Genetics:
A:Gene: 40
C:Superfamily: varicella-zoster virus gene 38 protein

Query Match	52.5%	Score	42	DB	2	Length	529
Best Local Similarity	45.5%	Pred. No.	17				
Matches	5	Conservative	5	Mismatches	1	Indels	0
						Gaps	0

```
QY      4  DQMVNVQVPWP  14
          {:::|::|
Db      283 DEILVEEPPWP 293
```

```

RESULT 13
C86468
Probable auxin response factor, 53188-50111 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86468
R:Theologis, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
C:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

```

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malti, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86468
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-600 <STO>
A:Cross-references: GB:AE005172; NID:g10092376; PIDN:AA612783.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match	52.5%	Score 42	DB 2	Length 600
Best Local Similarity	53.8%	Pred. No. 20		
Matches 7; Conservative	2	Mismatches 4	Indels 0	Gaps 0

QY 2 DQDQMVVVQVWP 14
| | : | : |
Db 566 DDDKMLVGDDPWP 578

```

RESULT 14
G86476
protein F1504.37 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

```

R.Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Aloni, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,ansen, N.F.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000
A.Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzetta, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A.Authors: Salberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallio,ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A.Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A.Reference number: A86141; MUID:21016719; PMID:11130712
A.Accession: G66476
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-767 <STO>
A.Cross-references: GB:A005172; NID:98778363; PIDN:AAF9371.1; GSPDB:CN00141
C.Genetics:
A.Gene: F1504.37
A.Map position: 1

Query Match	52.5%	Score 42:	DB 2;	Length 767;
Best Local Similarity	46.2%	Pred. NO:	26;	
Matches	6;	Mismatches	2;	Indels 0;
		Conservative		Gaps 0;

```
QY      2 DQDQMNVVQVPWP 14
          ::|||:|:|
Db      621 EEDKMLVGEDPWP 633
```

```

RESULT 15
W2BED4
gene 40 protein - equine herpesvirus 1 (strain Ab4p)
C:Species: equine herpesvirus 1
A:Note: host Equus caballus (domestic horse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C:Accession: F36799
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Submitted to Genbank, March 1992
A:Description: The DNA sequence of equine herpesvirus-1.
A:Reference number: A36805
A:Accession: F36799
A:Molecule type: DNA
A:Residues: 1-530 <TEL>
A:Cross-references: GB:M86664, NID:g330791, PIDN:AAB02476.1, PID:g330833

```

R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
 virology 189, 304-316, 1992
 A:Title: The DNA sequence of equine herpesvirus-1.
 A:Reference number: A41831; MUID:92295566; PMID:1318606
 A:Contents: annotation; possible protein-coding frames
 A:Note: neither amino acid nor nucleotide sequence is given
 C:Genetics:
 A:Gene: 40
 C:Superfamily: varicella-zoster virus gene 38 protein

Query Match 51.28; Score 41; DB 1; Length 530;
 Best Local Similarity 36.48; Pred. No. 26;
 Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Oy 4 DOMVVVQVPMP 14
 I: : : : :
 Db 284 DEILIEEPMP 294

Search completed: March 27, 2003, 10:06:52
 Job time : 6.0122 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 2.30488 Seconds
(without alignments)
251.930 Million cell updates/sec

Title: US-09-675-208-3

Perfect score: 80

Sequence: 1 YDDQWVYGVWPP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	55.0	229	1 AXIH_ARATH	P93830 arabidopsis
2	43	53.8	863	1 LQVR_CHICK	P98165 gallus gall
3	42	52.5	500	1 FQTB_ARATH	O9X178 arabidopsis
4	41	51.2	530	1 UL21_HSVB	P28972 equine herp
5	40	50.0	189	1 AXI6_ARATH	O38824 arabidopsis
6	40	50.0	653	1 ACSA_SYNY3	O54404 synecocyst
7	39	48.8	281	1 PHN2_ECOLI	P16688 escherichia
8	38	47.5	188	1 YECW_ECOLI	P52007 escherichia
9	38	47.5	404	1 WBS2_MOUSE	O54929 mus musculu
10	38	47.5	528	1 DYR2_HUMAN	O92630 mus sapien
11	38	47.5	574	1 ERG1_HUMAN	O14534 homo sapien
12	37	46.2	168	1 AXI1_ARATH	P49677 arabidopsis
13	37	46.2	174	1 AXI2_ARATH	P49678 arabidopsis
14	37	46.2	179	1 AXI6_PEA	P48680 pisum sativ
15	37	46.2	186	1 AXI4_ARATH	P33077 arabidopsis
16	37	46.2	188	1 AX2C_PHAU	O24541 phaseolus a
17	37	46.2	189	1 AXI4_PEA	O24542 phaseolus a
18	37	46.2	193	1 AX2D_PHAU	P49679 pisum sativ
19	37	46.2	194	1 AX2A_PHAU	P33293 phaseolus a
20	37	46.2	195	1 AX22_SOYB	P13088 glycine max
21	37	46.2	196	1 AX2B_PHAU	P33294 phaseolus a
22	37	46.2	203	1 AX2E_PHAU	O24543 phaseolus a
23	37	46.2	236	1 AXIG_ARATH	O24407 arabidopsis
24	37	46.2	243	1 AXI7_ARATH	O38825 glycine max
25	37	46.2	243	1 AXI7_ARATH	O38825 arabidopsis
26	37	46.2	251	1 SAST_ANAPL	P00633 anas platyr
27	37	46.2	261	1 AXIA_ARATH	O38826 arabidopsis
28	37	46.2	321	1 AXI8_ARATH	O38827 arabidopsis
29	37	46.2	338	1 AXI9_ARATH	GSP4027 xanthomonas
30	37	46.2	373	1 GSN2_XANCP	P14655 oryza sativ
31	37	46.2	428	1 GIN2_ORYSA	Oxhes0 mentha pipe
32	37	46.2	475	1 DXR_MENPI	Oxhes0 mentha pipe
33	37	46.2	477	1 DXR_ARATH	O9X179 arabidopsis

34	37	46.2	532	1 GRB7_HUMAN	O14451 homo sapien
35	37	46.2	709	1 ANDR_RABIT	P49699 oryctolagus
36	37	46.2	873	1 LQVR_RABIT	P35953 oryctolagus
37	37	46.2	884	1 ANDR_EULFC	O97776 eulimur ful
38	37	46.2	895	1 ANDR_MACRA	O97952 macaca fasc
39	37	46.2	895	1 ANDR_PAPHA	O97960 papio hamad
40	37	46.2	899	1 ANDR_MOUSE	P15091 mus musculu
41	37	46.2	902	1 ANDR_RAT	P15207 rattus norv
42	37	46.2	907	1 ANDR_CANFA	O97775 canis famli
43	37	46.2	911	1 ANDR_PANTR	O97775 pan troglod
44	37	46.2	919	1 ANDR_HUMAN	P10275 homo sapien
45	37	46.2	1006	1 BGAL_LACDE	P20043 lactobacill

ALIGNMENTS

RESULT 1
AXIH_ARATH STANDARD: PRT: 229 AA.
AC P93830: 049162:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Auxin-responsive protein IAA17 (Indoleacetic acid-induced protein 17)
DE (Auxin response 3).
GN IAA17 OR AXR3 OR ATIG04250 OR P19P19.31.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96004476; PubMed=9342315;
RA Kim J., Harter K., Theologis A.;
RT "Protein-protein interactions among the Aux/IAA proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:11786-11791(1997).
RN [2]
RP SEQUENCE FROM N.A., MUTANTS AXR3-1 AND AXR3-3, AND MUTAGENESIS OF
RP LEU-16; ASP-118 AND PRO-121.
RC STRAIN=cv. Columbia;
RX MEDLINE=98146427; PubMed=9478901;
RA Rouse D., Mackay P., Strinberg P., Estelle M., Leyser O.;
RT "Changes in auxin response from mutations in an AUX/IAA gene.";
RL Science 279:1371-1373(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egtu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
RA Utecher T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;

RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Toriumi M.J., Yu G., Brooks S.Y., Chao Q.,
 RA Chen H., Karlin-Neumann G., Kim C.J., Lam B., Miranda M., Nguyen M.,
 RA Palm C.J., Shinn P., Southwick A.M., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Full Length cDNA of gene F19P19.31 (GI:4389514).";
 RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT RIEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGEC).";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 CC GROWTH (BY SIMILARITY).
 CC -1- SUBUNIT: Homo and heterodimers. Interacts with the auxin response
 CC factors ARF1 and IAA24.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- INDUCTION: BY AUXIN.
 CC -1- MISCELLANEOUS: Increased auxin response of mutants aux3-1 and
 CC aux3-3 may result form an increased stability of AXR3.
 CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U49073; AAB84354.1; -;
 DR EMBL: AF040631; AAC39439.1; -;
 DR EMBL: AF040632; AAC39440.1; -;
 DR EMBL: AC000104; AAB70451.2; -;
 DR EMBL: AF36916; AAG53997.1; -;
 DR EMBL: AY070094; AAL49831.1; -;
 DR TRANSFAC: T04501; -;
 DR InterPro: IPR003311; AUX_IAA.
 DR Pfam: PF02309; AUX_IAA.1.
 DR MultiGene Family: Nuclear protein; Translation regulation.
 FT MUTAGEN 88 P->L: IN AXR3-1; INCREASES AUXIN
 FT RESPONSE.
 FT V->G: IN AXR3-3; INCREASES AUXIN
 FT RESPONSE.
 FT MUTAGEN 89 89 L->F: SUPPRESSES THE AXR3-1 PHENOTYPE;
 FT WHEN ASSOCIATED WITH L-88.
 FT MUTAGEN 118 118 D->N: SUPPRESSES THE AXR3-1 PHENOTYPE;
 FT WHEN ASSOCIATED WITH L-88.
 FT MUTAGEN 121 121 P->S: SUPPRESSES THE AXR3-1 PHENOTYPE;
 FT WHEN ASSOCIATED WITH L-88.
 SQ SEQUENCE 229 AA: 25288 MW: ECEA39207476581E CRC64;
 Query Match 55.0%; Score 44; DB 1; Length 229;
 Best Local Similarity 61.5%; Pred. No. 1.1;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 DDDQVVVVVYVWPP 14
 DB 185 DGDMMLVGDVWPP 197
 RESULT 2
 ID LDVR_CHICK STANDARD; PRT; 863 AA.
 AC P98165;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Very low-density lipoprotein precursor (VLDL receptor)
 DE (Vitellogenin receptor) (VNG receptor).

GN VLDLR OR VNGR.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn; TISSUE=ovary;
 RX MEDLINE=95045409; PubMed=7957081;
 RA Bujo H., Hermann M., Kaderli M.O., Jacobsen L., Sugawara S.,
 RA Nimpt J., Yamamoto T., Schneider W.J.;
 RT "Chicken oocyte growth is mediated by an eight ligand binding repeat
 RT member of the LDL receptor family.";
 RL EMBO J. 13:5165-5175(1994).
 RN [2]
 RP SEQUENCE OF 510-518; 546-554 AND 819-827.
 RC STRAIN=White leghorn; TISSUE=follicle membrane;
 RX MEDLINE=92011638; PubMed=1655760;
 RA Barber D.L., Sanders E.J., Abersold R., Schneider W.J.;
 RT "The receptor for yolk lipoprotein deposition in the chicken oocyte.";
 RL J. Biol. Chem. 266:18761-18770(1991).
 CC -1- FUNCTION: Binds VLDL and transports it into cells by endocytosis.
 CC In order to be internalized, the receptor-ligand complexes must
 CC first cluster into clathrin-coated pits. Binding to Reelin induces
 CC tyrosine phosphorylation of Dab1 and modulation of Tau
 CC phosphorylation (By similarity).
 CC -1- SUBUNIT: Binds to the extracellular matrix protein Reelin (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: ABUNDANT IN OOCYTES; MUCH LESS IN HEART AND
 CC SKELETAL MUSCLE.
 CC -1- SIMILARITY: CONTAINS 6 LDL-RECEPTOR CLASS B DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X80207; CAA56505.1; -;
 DR HSP: P01130; IAUJ.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_CA.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR000033; LDL_recept_rep.
 DR Pfam: PF00057; ldl_recept_a; 8
 DR Pfam: PF00058; ldl_recept_b; 5.
 DR PRINTS: PRO0261; LDLRECEPTOR.
 DR SMART: SM00179; EGF_CA; 2.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00192; LDLA; 8.
 DR SMART: SM00135; LY; 5.
 DR PROSITE: PS00010; ASX_HYDROXYL; 2.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS01187; EGF_CA; 2.
 DR PROSITE: PS01209; LDLRA_1; 8.
 DR PROSITE: PS50068; LDLRA_2; 8.
 DR Glycoprotein; VLDL; Cholesterol metabolism; Lipid transport;
 KW Endocytosis; Coated pits; Transmembrane; Receptor; Signal;
 KW EGF-like domain; Repeat.
 FT SIGNAL 1 43 POTENTIAL.
 FT CHAIN 44 863 VERY LOW-DENSITY LIPOPROTEIN RECEPTOR.
 FT DOMAIN 44 785 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 786 809 POTENTIAL.
 FT DOMAIN 810 863 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 49 87 LDL-RECEPTOR CLASS A 1.


```

FT DOMAIN 88 128 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 129 169 LDL-RECEPTOR CLASS A 3.
FT DOMAIN 170 208 LDL-RECEPTOR CLASS A 4.
FT DOMAIN 209 249 LDL-RECEPTOR CLASS A 5.
FT DOMAIN 255 293 LDL-RECEPTOR CLASS A 6.
FT DOMAIN 294 332 LDL-RECEPTOR CLASS A 7.
FT DOMAIN 334 373 LDL-RECEPTOR CLASS A 8.
FT DOMAIN 374 413 EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 414 453 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 457 498 LDL-RECEPTOR CLASS B 1.
FT REPEAT 499 544 LDL-RECEPTOR CLASS B 2.
FT REPEAT 545 587 LDL-RECEPTOR CLASS B 3.
FT REPEAT 588 631 LDL-RECEPTOR CLASS B 4.
FT REPEAT 632 674 LDL-RECEPTOR CLASS B 5.
FT REPEAT 675 716 LDL-RECEPTOR CLASS B 6.
FT DOMAIN 722 770 EGF-LIKE 3.
FT SITE 822 827 ENDOTOXIN SIGNAL (POTENTIAL).
FT DISULFID 51 63 BY SIMILARITY.
FT DISULFID 58 76 BY SIMILARITY.
FT DISULFID 70 85 BY SIMILARITY.
FT DISULFID 90 102 BY SIMILARITY.
FT DISULFID 97 115 BY SIMILARITY.
FT DISULFID 109 126 BY SIMILARITY.
FT DISULFID 131 145 BY SIMILARITY.
FT DISULFID 138 158 BY SIMILARITY.
FT DISULFID 152 167 BY SIMILARITY.
FT DISULFID 172 184 BY SIMILARITY.
FT DISULFID 179 197 BY SIMILARITY.
FT DISULFID 191 206 BY SIMILARITY.
FT DISULFID 211 223 BY SIMILARITY.
FT DISULFID 223 236 BY SIMILARITY.
FT DISULFID 230 247 BY SIMILARITY.
FT DISULFID 257 269 BY SIMILARITY.
FT DISULFID 264 282 BY SIMILARITY.
FT DISULFID 276 291 BY SIMILARITY.
FT DISULFID 296 308 BY SIMILARITY.
FT DISULFID 303 321 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 336 349 BY SIMILARITY.
FT DISULFID 344 362 BY SIMILARITY.
FT DISULFID 356 373 BY SIMILARITY.
FT DISULFID 378 389 BY SIMILARITY.
FT DISULFID 385 398 BY SIMILARITY.
FT DISULFID 400 412 BY SIMILARITY.
FT DISULFID 418 428 BY SIMILARITY.
FT DISULFID 424 437 BY SIMILARITY.
FT DISULFID 439 452 BY SIMILARITY.
FT DISULFID 726 739 BY SIMILARITY.
FT DISULFID 735 754 BY SIMILARITY.
FT CARBOHYD 756 769 BY SIMILARITY.
FT CARBOHYD 169 169 N-LINKED (GLUCAN. . .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLUCAN. . .) (POTENTIAL).
SO SEQUENCE 863 AA: 94904 MW: 0672AB74BF9A2245 CRC64;

```

```

Query Match 53.8%; Score 43; DB 1; Length 863;
Best Local Similarity 35.7%; Pred. No. 7.1;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

```

Oy 1 YDQDMVVVQVWP 14
Db 609 FDRQDLVTEIOWP 622

```

RESULT 3

```

FUT8_ARATH STANDARD: PRT: 500 AA.

```

```

AC 05X178;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable fucosyltransferase 8 (EC 2.4.1.-) (AtFUT8).
GN FUT8 OR AT1G14100 OR F7A19.18.
OS Arabidopsis thaliana (Mouse-ear cress).

```

```

OC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA White O., Alonso J., Altabi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujil C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Niermen W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Rambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408:816-820(2000).
RN [2]
RP IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY.
RX MEDLINE=21608393; PubMed=11743104;
RA Sarría R., Wagner T.A., O'Neill M.A., Falk A., Wilkerson C.G.,
RA Keegstra K., Raikhel N.V.;
RT "Characterization of a family of Arabidopsis genes related to
xyloglucan fucosyltransferase1."
RL Plant Physiol. 127:1595-1606(2001).
CC - FUNCTION: May be involved in cell wall biosynthesis. May
act as a fucosyltransferase.
CC - PATHWAY: Glycosylation.
CC - SUBCELLULAR LOCATION: Golgi (Potential).
CC - TISSUE SPECIFICITY: Expressed in leaves and stems.
CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AC007576; AAC39294.1; -
DR InterPro: IPR004938; XG_Ftase.
DR Pfam: PF03254; XG_Ftase; 1
KW ftransferase; Glycosyltransferase; Golgi stack; Cell wall.
SO SEQUENCE 500 AA: 57538 MW: 1C5BB6152BFAE690 CRC64;

```

```

Query Match 52.5%; Score 42; DB 1; Length 500;
Best Local Similarity 50.0%; Pred. No. 6;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 2 DQDMVVVQVWP 13
Db 225 ESDQILIRQVWP 236

```

RESULT 4

```

UL21_HSVB STANDARD: PRT: 530 AA.

```

```

AC P28972;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Gene 40 protein.

```

```

GN 40.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1."
RL Virology 189:304-316(1992).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL21,
CC EHV-1 40, EHV-4 UL21, AND VZV 38.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M86664; AAB02476.1; -.
DR PIR: F36799; WZBBD4.
DR InterPro: IPR004936; Herpes_UL21.
DR Pfam: PF03252; UL21; 1.
SQ SEQUENCE 530 AA; 57915 MW; 2E43565AB297CD18 CRC64;

Query Match
Best Local Similarity 36.4%; Pred. No. 9.7; Length 530;
Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 4 DQMVVVQVMP 14
ID 1:1:1:1:1
DB 284 DEILLIEPMP 294

RESULT 5
AA16_ARATH STANDARD: PRT; 189 AA.
ID AX16_ARATH OGC539;
AC 038824;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Auxin-responsive protein IAA6 (Indoleacetic acid-induced protein 6).
GN IAA6 OR AIG52830 OR FI4G24.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RA MEDLINE=95387393; PubMed=7658471;
RA Abel S., Nguyen M.D., Theologis A.;
RT "The PS-1AA4/5-like family of early auxin-inducible mRNAs in
RT Arabidopsis thaliana."
RL J. Mol. Biol. 251:533-549(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RA MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
RA Dunn P., Elgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huitzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marshall A.,

```

```

RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Unterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Toriumi M., Yu G., Brooks S., Chao Q.,
RA Chen H., Karlin-Neumann G., Kim C., Lam B., Miranda M., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene FI4G24.10 (GI:6862944)."
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
CC GROWTH (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18408; AAC49047.1; -.
DR EMBL: AC019018; AAG52268.1; -.
DR EMBL: AF336915; AAG53996.1; -.
DR InterPro: IPR003311; AUX_IAA.
DR Pfam: PF02309; AUX_IAA; 1.
KW Multigene family; Nuclear protein; translation regulation.
FT CONFIRM 34 34 K -> N (IN REF. 1).
SQ SEQUENCE 189 AA; 21031 MW; 5EB9130584A75465 CRC64;

Query Match
Best Local Similarity 50.0%; Score 40; DB 1; Length 189;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 DQDVVVQVMP 13
ID 1:1:1:1:1
DB 152 DRDMLVGVMP 163

RESULT 6
ACSA_SYNY3 STANDARD: PRT; 653 AA.
ID ID ACSA_SYNY3
AC 055404;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A synthetase (EC 6.2.1.1) (acetate--CoA ligase) (Acy1-
DE activating enzyme).
GN ACS OR SL0542.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 648 to 928 of the genome."
RL DNA Res. 2:153-166(1995).

```

```

CC -1- CATALYTIC ACTIVITY: ATP + acetate + CoA -> AMP + diphosphate +
CC acetyl-CoA.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64003; BA10498.1; -.
DR HSSP; P08659; 1LC1.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding.1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00455; AMP_BINDING.1.
KW L19aag; Complete proteome.
SQ SEQUENCE 653 AA; 73052 MW; 5EBCAD430EEFFA03 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 653;
Best Local Similarity 46.2%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 DQDMVVQVWP 14
   11 :11:111
Db 464 DQGFVITKQPP 476

RESULT 7
ID PHNJ_ECOLI STANDARD: PRT: 281 AA.
AC P16688;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PhnJ protein.
GN PHNJ OR B4098.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12.
RX MEDLINE=91193228; PubMed=1840580;
RA Makino K., Kim S.K., Shingawa H., Amemura M., Nakata A.;
RT "Molecular analysis of the cryptic and functional phn operons for
RT phosphonate use in Escherichia coli K-12.";
RL Bacteriol. 173:2655-2672(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=B;
RX MEDLINE=90170953; PubMed=2155230;
RA Chen C.-M., Ye Q.-Z., Zhu Z., Warner B.L., Walsh C.T.;
RT "Molecular biology of carbon-phosphorus bond cleavage. Cloning and
RT sequencing of the phn (pslD) genes involved in alkyldiphosphonate
RT uptake and C-P lyase activity in Escherichia coli B.";
RL J. Biol. Chem. 265:4461-4471(1990).
RN [4]
RP FUNCTION: BELONGS TO AN OPERON INVOLVED IN ALKYLPHOSPHONATE
CC UPTAKE AND C-P LYASE. EXACT FUNCTION NOT KNOWN.
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN K12.

```

```

CC -1- SIMILARITY: BELONGS TO THE PHNJ FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90227; BA114270.1; -.
DR EMBL; U14003; AA96997.1; -.
DR EMBL; AE000482; AAC77059.1; -.
DR EMBL; J05260; AAA24348.1; -.
DR PIR; B35719; B35719.
DR Ecocore; EG10719; phnJ.
KW Alkyldiphosphonate uptake; Complete proteome.
FT VARIANT 103 103 V->L (IN STRAIN B).
SQ SEQUENCE 281 AA; 31845 MW; 241F6AF14095468 CRC64;

Query Match 48.8%; Score 39; DB 1; Length 281;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 QDMVVQVWP 14
   11 :11:111
Db 117 EDQIIFQVPIP 128

RESULT 8
ID YECM_ECOLI STANDARD: PRT: 188 AA.
AC P52007; P52008; P76295;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein yecM.
GN YECM OR B1875.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshimi H.,
RA Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
RN [3]
RP SEQUENCE OF 102-188 FROM N.A.
RC STRAIN=K12 / MC4100;
RX Gupta S.D., Wu H.C.;
RT Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=99420866; PubMed=10493123;

```

```

RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography."
RT Electrophoresis 20:2181-2195(1999).
CC -1- SIMILARITY: STRONG, TO H.INFLUENZAE H11582/H11581.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 128 THAT PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000281; AAC74945.1; ALT_INIT.
CC EMBL: D90829; BAA15685.1; -.
CC EMBL: D90830; BAA15692.1; -.
CC EMBL: J38618; AAA89200.1; ALT_FRAME.
CC EMBL: J38618; AAA89201.1; ALT_FRAME.
CC Ecogene; EG13405; yecm.
CC Complete proteome.
CC CONFLICT 102 102 G -> R (IN REF. 3).
CC CONFLICT 120 120 G -> V (IN REF. 3).
CC CONFLICT 167 167 G -> A (IN REF. 3).
CC FT SEQUENCE 188 AA; 21205 MW; 700B8E042F8AA8A5 CRC64;
SQ
Query Match 47.5%; Score 38; DB 1; Length 188;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 5 QMYYVQVPWP 14
DB 92 QFSIVELPWP 101

```

RESULT 9

```

WBS2_MOUSE
ID WBS2_MOUSE STANDARD; PRT; 404 AA.
AC 054929; O9R1M8;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE WD repeat and SOCS box containing protein 2 (WSB-2) (SOCS box-
DE containing WD protein SWIP-2).
GN WSB2 OR SWIP2.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
CC NCBI_TaxID=10090;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=C57BL/6; TISSUE=Spleen;
CC MEDLINE=98081836; PubMed=9419338;
CC RA Hilton D.J., Richardson R.T., Alexander W.S., Viney E.M.,
CC Wilson T.A., Sprigg N.S., Starr R., Nicholson S.E., Metcalf D.,
CC Nicola N.A.;
CC "Twenty proteins containing a C-terminal SOCS box form five structural
CC classes."
CC Proc. Natl. Acad. Sci. U.S.A. 95:114-119(1998).
CC [2]
CC SEQUENCE FROM N.A.
CC MEDLINE=99284520; PubMed=10354473;
CC RA Vasilakos D., Hancock S., Stern C.D.;
CC "SWIP-1: novel SOCS box containing WD-protein regulated by signalling
CC centres and by Shb during development."
CC Mech. Dev. 82:79-94(1999).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: CONTAINS 1 SOCS BOX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF03188; AAB96649.1; -.
CC EMBL: AF072881; AAD28809.1; -.
CC InterPro: IPR001496; SOCS.
CC InterPro: IPR001680; WDA0.
CC Pfam: PF00400; WDA0; 6.
CC PRINTS; PR00320; GPROTEINBPT.
CC PRODOM; PD000018; WDA0; 2.
CC SMART; SM00253; SOCS; 1.
CC SMART; SM00320; WDA0; 6.
CC PROSITE; PS00678; WD_REPEATS_1; 2.
CC PROSITE; PS50082; WD_REPEATS_2; 5.
CC PROSITE; PS50284; WD_REPEATS_REGION; 1.
CC PROSITE; PS50225; SOCS; 1.
CC Repeat; WD repeat.
CC FT REPEAT 105 148 WD 1.
CC FT REPEAT 151 191 WD 2.
CC FT REPEAT 195 234 WD 3.
CC FT REPEAT 237 276 WD 4.
CC FT REPEAT 291 330 WD 5.
CC FT DOMAIN 356 404 SOCS BOX.
CC CONFLICT 272 272 A -> E (IN REF. 2).
CC FT SEQUENCE 404 AA; 45203 MW; 78F66AF49AC4659A CRC64;
SQ
Query Match 47.5%; Score 38; DB 1; Length 404;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
OY 1 YDDQMWVWVPWP 14
DB 43 WSGGHCYVKLVWPWP 56

```

RESULT 10

```

DYR2_HUMAN
ID DYR2_HUMAN STANDARD; PRT; 528 AA.
AC 092630;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 2
DE (EC 2.7.1.-).
GN DYRK2.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC NCBI_TaxID=9606;
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Brain;
CC MEDLINE=98421512; PubMed=9748265;
CC RA Becker W., Weber Y., Wetzel K., Erimbler K., Tejedor F.J.,
CC Joost H.-G.;
CC "Sequence characteristics, subcellular localization, and substrate
CC specificity of DYRK-related kinases, a novel family of dual
CC specificity protein kinases."
CC J. Biol. Chem. 273:25893-25902(1998).
CC [2]
CC SEQUENCE OF 320-528 FROM N.A.
CC TISSUE=Placenta;
CC RA Becker W., Joost H.-G.;
CC Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN VITRO; CAN PHOSPHORYLATE HISTONES H3 AND H2B ON SER
CC AND THR RESIDUES. MAY BE INVOLVED IN THE REGULATION OF CELLULAR
CC GROWTH AND/OR DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

```

CC MNB/DYRK SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y13493; CA73885.1; -
 CC EMBL: Y09216; CA70418.1; -
 CC HSSP: Q00534; 1B18.
 CC Genew: HGNC:3093; DYRK2.
 CC MIM: 603496; -
 CC InterPro: IPR000719; Euk_pkinase.
 CC InterPro: IPR002290; Ser_thr_pkinase.
 CC Pfam: PF00069; pkinase; 1.
 CC ProDom: PD000001; Euk_pkinase; 1.
 CC SMART: SM00220; S_TKC; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 CC Transferrase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
 CC ATP-binding; Phosphorylation.
 CC KW DOMAIN 149 462
 CC NP_BIND 155 163 ATP (BY SIMILARITY).
 CC BINDING 178 178 ATP (BY SIMILARITY).
 CC ACT_SITE 275 275 BY SIMILARITY.
 CC SO SEQUENCE 528 AA; 59714 MW; AF2C682ED9522D7 CRC64;
 CC
 CC Query Match 47.5%; Score 38; DB 1; Length 528;
 CC Best Local Similarity 66.7%; Pred. No. 33;
 CC Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 CC
 CC 1 YDQDMVYVQVP 12
 CC 1111111111
 CC 130 YDDDCSYVQVP 141
 CC
 CC RESULT 11
 CC ERGL_HUMAN STANDARD; PRT; 574 AA.
 CC ID ERGL_HUMAN
 CC AC Q14534; Q9UEK6;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Squalene monooxygenase (EC 1.14.99.7) (Squalene epoxidase) (SE).
 CC GN SQUE OR ERGL.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_Taxid=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=97432831; PubMed=9286711;
 CC RA Nagai M., Sakakibara J., Wakui K., Fukushima Y., Igarashi S.,
 CC Tanji S., Arakawa M., Ono T.;
 CC "Localization of the squalene epoxidase gene (SQUE) to human
 CC chromosome region Bq24.1";
 CC RT Genomics 44:141-143(1997).
 CC RL [2]
 CC RN SEQUENCE OF 187-535 FROM N.A.
 CC RP TISSUE=Liver;
 CC RC MEDLINE=96215195; PubMed=8626488;
 CC RA Nakamura T., Sakakibara J., Izumi T., Shibata A., Ono T.;
 CC "Transcriptional regulation of squalene epoxidase by sterols and
 CC RT inhibitors in HeLa cells";
 CC J. Biol. Chem. 271:8053-8056(1996).
 CC -1- FUNCTION: CATALYZES THE FIRST OXYGENATION STEP IN STEROL
 CC BIOSYNTHESIS AND IS SUGGESTED TO BE ONE OF THE RATE-LIMITING
 CC ENZYMES IN THIS PATHWAY.
 CC -1- CATALYTIC ACTIVITY: Squalene + AH(2) + O(2) = (S)-squalene-2,3-

CC epoxide + A + H(2)O.
 CC -1- COFACTOR: FAD.
 CC -1- SUBUNIT: MAY FORM A COMPLEX WITH SQUALENE SYNTHASE.
 CC -1- SUBCELLULAR LOCATION: Mitosomal.
 CC -1- SIMILARITY: BELONGS TO THE SQUALENE MONOOXYGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D78130; BAA22372.1; -
 CC EMBL: D78129; BAA11209.1; -
 CC Genew: HGNC:11279; SQUE.
 CC MIM: 602019; -
 CC InterPro: IPR000733; Flav_monooxygenase.
 CC InterPro: IPR000205; NAD_binding.
 CC InterPro: IPR003042; Rog_monooxygenase.
 CC Pfam: PF01360; Monooxygenase; 1.
 CC PRINTS: PR00420; RINGMOXGNASE.
 CC Oxidoreductase; Flavoprotein; FAD; Transmembrane.
 CC KW TRANSMEM 20 40
 CC FT TRANSMEM 61 81 POTENTIAL.
 CC FT TRANSMEM 123 143 POTENTIAL.
 CC FT TRANSMEM 546 566 POTENTIAL.
 CC FT NP_BIND 126 153 FAD (ADP PART) (POTENTIAL).
 CC FT CONFLICT 247 247 E -> G (IN REF. 2).
 CC FT CONFLICT 336 336 R -> Q (IN REF. 2).
 CC FT CONFLICT 389 389 L -> P (IN REF. 2).
 CC FT CONFLICT 451 451 N -> K (IN REF. 2).
 CC FT CONFLICT 518 518 A -> V (IN REF. 2).
 CC SQ SEQUENCE 574 AA; 63939 MW; DDAF95308864FD09 CRC64;
 CC
 CC Query Match 47.5%; Score 38; DB 1; Length 574;
 CC Best Local Similarity 42.9%; Pred. No. 36;
 CC Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 CC
 CC 1 YDQDMVYVQVP 14
 CC 1111111111
 CC 198 HDQSKSEVQIDVP 211
 CC
 CC RESULT 12
 CC AXIL_ARATH STANDARD; PRT; 168 AA.
 CC ID AXIL_ARATH
 CC AC P49677; O23312;
 CC DT 01-FEB-1996 (Rel. 33, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Auxin-responsive protein IAA1 (Indoleacetic acid-induced protein 1).
 CC GN IAA1 OR A74G14560 OR D1J320W.
 CC OS Arabidopsis thaliana (Mouse-ear cress).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 CC NCBI_Taxid=3702;
 CC RN [1]
 CC RN SEQUENCE FROM N.A.
 CC RP STRAIN=cv. Columbia;
 CC RX MEDLINE=94105161; PubMed=8278386;
 CC RA Abel S., Oeller P.W., Theologis A.;
 CC "Early auxin-induced genes encode short-lived nuclear proteins";
 CC Proc. Natl. Acad. Sci. U.S.A. 91:326-330(1994).
 CC [2]
 CC RN SEQUENCE FROM N.A.
 CC RP STRAIN=cv. Columbia;
 CC RX MEDLINE=20083488; PubMed=10617198;
 CC RA Mayer K.F.X., Schueller C., Wambolt R., Murphy G., Volckaert G.,
 CC Pohl T., Duesterhoeft A., Stiekema W., Ertlan K.-D., Terry N.,
 CC Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,

RA Weichselgartner M., de Simone V., Obermaier B., Macho R., Mueller M.,
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmalthein T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hebeisel J., Zimmermann M., Wedler F., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grynoprez B., Chuang Y.-J., Vandebussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Deloor E.,
 RA Welteneberger T., Botte G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirker W.,
 RA Moellman P., Klein lankhorst R., Rose M., Hauf J., Koelter P.,
 RA Bernier S., Hempel S., Feldpausch M., Lambirth S., Van den Daele H.,
 RA De Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McElroy K., Mayes R.,
 RA Petlett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
 RA Borjova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Heitzl A.,
 RA Neumann S., Argitrou A., Vitale D., Liguori R., Pitravadi E.,
 RA Massenot O., Ouligley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chetdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker H.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhi N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoeckling T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latteille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shoudy N., Hasegawa A., Hamed A., Lohdl M., Johnson A.,
 RA Chen E., Maria M., Martienssen R., McCombie W.R.,
 RT *Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.
 RT Nature 402:769-777(1999).
 RL
 CC -!- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 CC GROWTH.
 CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 CC GROWTH.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- INDUCTION: BY AUXIN.
 CC -!- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, L15448; AAA16569.1; -;
 CC DR EMBL, 297336; CAB10235.1; -;
 CC DR EMBL, AL161539; CAB78498.1; -;
 CC DR TRANSFAC; T04508; -;
 CC DR InterPro; IPR003311; AUX_IAA.
 CC Pfam; PF02309; AUX_IAA; 1.
 CC MultiGene family: Nuclear protein; Translation regulation.
 CC KW CONFLICT 52 52 A -> T (IN REF. 1).
 CC FT SEQUENCE 168 AA; 19031 MW; 17CA9B3BBE7962EE CRC64;
 SO

DB 135 DGDMLVGPWP 146
 RESULT 13
 AX12_ARATH STANDARD; PRT; 174 AA.
 AC P49678; O22596;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Auxin-responsive protein IAA2 (indoleacetic acid-induced protein 2).
 GN IAA2 OR AT3G23030 OR MXC7.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=94105161; PubMed=8278386;
 RA Abel S., Oeller P.W., Theologis A.;
 RT "Early auxin-induced genes encode short-lived nuclear proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:326-330(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Wassilewskaja;
 RA Malik M.K., Zimmerman J.L.;
 RT "Molecular characterization of a genomic clone encoding the auxin-
 RT inducible IAA2 gene from Arabidopsis thaliana";
 RL (in) Plant Gene Register PGR97-178.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and YAC
 RT clones";
 RL DNA Res. 7:131-135(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Egu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene MXC7.6 (GI:9294192).";
 CC Submitted (Dec-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 CC GROWTH.
 CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 CC GROWTH.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- INDUCTION: BY AUXIN.
 CC -!- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, L15449; AAA16570.1; -;
 CC DR EMBL, AF027157; AAB97164.1; -;
 CC DR EMBL, AB026655; BAB02094.1; -;
 CC DR EMBL, AF332392; AAG48756.1; -;
 CC DR TRANSFAC; T04509; -;
 CC DR InterPro; IPR003311; AUX_IAA.
 CC Pfam; PF02309; AUX_IAA; 1.
 CC MultiGene family: Nuclear protein; Translation regulation.
 KW

FT CONFLICT 38 38 S -> I (IN REF. 2).
SQ SEQUENCE 174 AA; 19910 MW; DECA4F59F0106179 CRC64;
Query Match
Best Local Similarity 58.3%; Score 37; DB 1; Length 174;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 2 DODOMVVVQVWPV 13
| | | | |
Db 138 DGDWMLVGDVPM 149
RESULT 14
AX16_PEA STANDARD; PRT; 179 AA.
AC P49680;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Auxin-induced protein IAA6.
GN IAA6.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska;
RX MEDLINE=94016594; PubMed=8411182;
RA Oeller P.W., Keller J.A., Parks J.E., Silbert J.E., Theologis A.;
RT "Structural characterization of the early indoleacetic acid-inducible
genes, PS-IAA4/5 and PS-IAA6, of pea (Pisum sativum L.).";
RL J. Mol. Biol. 233:789-798(1993).
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
CELL GROWTH (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC DR EMBL: X68218; CAA48300.1; -;
CC DR EMBL: X68217; CAA48299.1; -;
CC DR InterPro: IPR003311; AUX_IAA.
CC Pfam: PF02309; AUX_IAA; 1.
CC MultiGene family: Nuclear protein; Translation regulation.
CC SO SEQUENCE 179 AA; 20330 MW; 0D212F384D100D7E CRC64;
Query Match
Best Local Similarity 58.3%; Score 37; DB 1; Length 179;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 2 DODOMVVVQVWPV 13
| | | | |
Db 135 DGDWMLVGDVPM 146
RESULT 15
AX14_ARATH STANDARD; PRT; 186 AA.
AC P33077;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Auxin-responsive protein IAA4 (indoleacetic acid-induced protein 4)

DE (Auxin-induced protein AUX2-11).
GN IAA4 OR AUX2-11 OR AT5G43700 OR MOD19.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=91338700; PubMed=2102379;
RA Conner T.W., Goekjian V.H., Lafayette P.R., Key J.L.;
RT "Structure and expression of two auxin-inducible genes from
RT Arabidopsis.";
RL Plant Mol. Biol. 15:623-632(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=94105161; PubMed=8278386;
RA Abel S., Oeller P.W., Theologis A.;
RT "Early auxin-induced genes encode short-lived nuclear proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:326-330(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RT Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Yamada K., Liu S.X., Sakano H., Phan P.K., Banh J., Etyu P., Lee J.M.,
RA Tortum K., Yu G., Brooks S., Chao Q., Chen H., Katlin-Neumann G.,
RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MOD19.3 (GI:10177938).";
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
CELL GROWTH.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC DR EMBL: X53435; CAA37526.1; -;
CC DR EMBL: L15450; AAA16571.1; -;
CC DR EMBL: AB026651; BAB11297.1; -;
CC DR EMBL: AF332394; AAC48758.1; -;
CC DR PIR: S12243; S12243.
CC DR TRANSFAC: T04511; -;
CC DR InterPro: IPR003311; AUX_IAA.
CC Pfam: PF02309; AUX_IAA; 1.
CC MultiGene family: Nuclear protein; Translation regulation.
CC SO SEQUENCE 186 AA; 20975 MW; DCF66CAC4D28CD3B CRC64;
Query Match
Best Local Similarity 58.3%; Score 37; DB 1; Length 186;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 2 DODOMVVVQVWPV 13
| | | | |
Db 149 DGDWMLVGDVPM 160

Fri Apr 4 12:31:26 2003

us-09-675-208-3.rsp

Page 10

Search completed: March 27, 2003, 10:04:18
Job time : 3.30488 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 7.85366 Seconds
(without alignments)
367.301 Million cell updates/sec

Title: US-09-675-208-3
Perfect score: 80
Sequence: 1 YDQDMVVGVFWP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	100.0	784	10 Q92T62	Q92T62 cucumis sat
2	49	61.3	783	10 Q9FND9	Q9FND9 arabidopsis
3	45	56.2	570	10 Q9L0E3	Q9L0E3 arabidopsis
4	44	55.0	164	10 Q38832	Q38832 arabidopsis
5	44	55.0	228	10 Q9C5W8	Q9C5W8 arabidopsis
6	44	55.0	234	10 Q23311	Q23311 arabidopsis
7	44	55.0	619	10 Q9X1D4	Q9X1D4 arabidopsis
8	44	55.0	719	10 Q9LXN1	Q9LXN1 arabidopsis
9	44	53.8	416	16 Q9B1L8	Q9B1L8 rhizobium 1
10	43	53.8	418	16 Q8U0J3	Q8U0J3 agrobacteri
11	43	53.8	423	10 Q81350	Q81350 enteromorph
12	42	52.5	462	10 Q8S385	Q8S385 phaseolus 1
13	42	52.5	529	12 Q39281	Q39281 equine herp
14	42	52.5	600	10 Q9C8N7	Q9C8N7 arabidopsis
15	42	52.5	767	10 Q9L0E8	Q9L0E8 arabidopsis
16	42	52.5	798	10 Q8VWN6	Q8VWN6 pisum sativ

17	40	50.0	202	10 Q9XEX3	Q9XEX3 nicotiana t
18	40	50.0	346	16 Q8YUM9	Q8YUM9 anabena sp
19	40	50.0	448	16 Q97SM0	Q97SM0 streptococ
20	40	50.0	847	2 Q9ADP8	Q9ADP8 salmonella
21	40	50.0	1067	2 Q9NN63	Q9NN63 leishmania
22	40	50.0	1159	4 Q9UF39	Q9UF39 homo sapien
23	39.5	49.4	264	16 Q988P9	Q988P9 rhizobium 1
24	39.5	49.4	345	7 Q19480	Q19480 heterodoncu
25	39	48.8	93	5 Q9U194	Q9U194 leishmania
26	39	48.8	206	5 Q00629	Q00629 manduca sex
27	39	48.8	281	2 Q93HT2	Q93HT2 bradyrhizob
28	39	48.8	281	16 Q8X5P0	Q8X5P0 escherichia
29	39	48.8	288	5 Q9XZD6	Q9XZD6 plasmodium
30	39	48.8	305	10 Q9ZSY8	Q9ZSY8 arabidopsis
31	39	48.8	310	17 Q96XN0	Q96XN0 sulfolobus
32	39	48.8	328	3 Q9CUB5	Q9CUB5 neurospora
33	39	48.8	384	16 Q9PD55	Q9PD55 xyella fas
34	39	48.8	565	12 Q9PY76	Q9PY76 human calic
35	39	48.8	566	12 Q8VA19	Q8VA19 human calic
36	39	48.8	598	5 Q963J1	Q963J1 drosophila
37	39	48.8	848	13 Q9XGV9	Q9XGV9 anguilla ja
38	39	48.8	2435	5 Q9W433	Q9W433 drosophila
39	38.5	48.1	190	8 Q94Y27	Q94Y27 pyralisella 1
40	38	47.5	73	4 Q9HC65	Q9HC65 homo sapien
41	38	47.5	138	16 Q9FCE9	Q9FCE9 streptomyce
42	38	47.5	162	16 Q9PCF0	Q9PCF0 xyella fas
43	38	47.5	175	10 Q9XEL4	Q9XEL4 oryza sativ
44	38	47.5	190	16 Q8XCH3	Q8XCH3 escherichia
45	38	47.5	281	16 Q8Y219	Q8Y219 raistonia s

ALIGNMENTS

RESULT 1					
Q92T62		PRELIMINARY:	PRT:	784 AA.	
AC Q92T62;					
DT 01-MAY-1999 (TREMBLrel. 10, Created)					
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)					
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)					
DE Raffinose synthase (EC 2.4.1.82).					
GN RFS.					
OS Cucumis sativus (Cucumber).					
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:					
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:					
OC Cucurbitales: Cucurbitaceae: Cucumis.					
OX NCBI_TaxID=3659;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC TISSUE=LEAF:					
RA Ohsumi C., Nozaki J., Kida T.:					
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.					
DR EMBL: AF073744; AAD02832.1; -					
KW Glycosyltransferase; Transferase.					
SQ SEQUENCE 784 AA: 86920 MW: 38064491F0908933 CRC64:					
Query Match	100.0%:	Score 80:	DB 10:	Length 784:	
Best Local Similarity	100.0%:	Pred. No. 6.4e-06:			
Matches 14:	Conservative	0:	Mismatches 0:	Indels 0:	Gaps 0:
OY 1 YDQDMVVGVFWP 14					
DB 756 YDQDMVVGVFWP 769					
RESULT 2					
Q9FND9		PRELIMINARY:	PRT:	783 AA.	
ID Q9FND9;					
DT 01-MAR-2001 (TREMBLrel. 16, Created)					
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)					
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)					

```

DE Rafinose synthase protein.
GN Mp012.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RX (1)
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones."
RL DNA Res. 4:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB006702; BAB11595.1; -;
DR EMBL: AY062781; AAL32859.1; -;
DR EMBL: AY081645; AAM1207.1; -;
SO SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;

Query Match 61.3%; Score 49; DB 10; Length 783;
Best Local Similarity 72.7%; Pred. No. 2.6;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 DODMWWGVVWPP 13
DB 760 EDSMWWGVVWPP 770

RESULT 3
Q9LOE3 PRELIMINARY; PRT; 570 AA.
AC Q9LOE3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE P1504.42.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RX (1)
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,

```

```

RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chiu J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharzky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thavert A., Tortum M., Vaysberg M., Yu C., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007887; AAF79360.1; -;
DR InterPro: IPR003340; TF_B3.
DR Pfam: PF02362; B3; 1.
SO SEQUENCE 570 AA; 65041 MW; E64A987BA713DF7E CRC64;

Query Match 56.2%; Score 45; DB 10; Length 570;
Best Local Similarity 53.8%; Pred. No. 9.9;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 DODMWWGVVWPP 14
DB 558 DEDEMVLGDDPWP 570

RESULT 4
Q38832 PRELIMINARY; PRT; 164 AA.
AC Q38832;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE IAA14 (Fragment).
GN IAA14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RX (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=95387393; PubMed=7658471;
RA Abel S., Nguyen M.D., Theologis A.;
RT "The PS-IAA4/5-like family of early auxin-inducible mRNAs in
RT Arabidopsis thaliana."
RL J. Mol. Biol. 251:533-549(1995).
DR EMBL: U18416; AAC49055.1; -;
DR InterPro: IPR003311; AUX_IAA.
DR Pfam: PF02309; AUX_IAA; 1.
DR NON_TER 1
FT 1
SO SEQUENCE 164 AA; 18232 MW; D7BB6B71F6DD3574 CRC64;

Query Match 55.0%; Score 44; DB 10; Length 164;
Best Local Similarity 61.5%; Pred. No. 3.9;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DODMWWGVVWPP 14
DB 120 DGDMLVGVDPWP 132

RESULT 5
Q9C5W8 PRELIMINARY; PRT; 228 AA.
AC Q9C5W8;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE IAA14.
GN D13315C/ATG14550.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RX (1)
RP SEQUENCE FROM N.A.

```

RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Toriumi M., Yu G., Brooks S., Chao Q.,
 RA Chen H., Karlin-Neumann G., Kim C., Lam B., Miranda M., Nguyen M.,
 RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Full Length cDNA of gene d13315c/At4g14550 (GI:7268161).";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF3474718; AAC50096.1; -
 DR InterPro: IPR003311; AUX_1AA.
 DR Pfam: PF02309; AUX_1AA; 1.
 DR SEQUENCE 228 AA; 2504 MW; A76A9F733538E372 CRC64;
 SO

Query Match 55.0%; Score 44; DB 10; Length 228;
 Best Local Similarity 61.5%; Pred. No. 5.6;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 2 DODOMVVQVWMP 14
 Db 184 DGDWMLVGDVWMP 196

RESULT 6

ID 023311 PRELIMINARY; PRT; 234 AA.
 AC 023311;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE IAA7-like protein (IAA7 like protein).
 GN DLJ315C OR AT4G14550.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Stiekema W., Murphy G., Mambutt R., Pohl T., Terry N.,
 RA Kreis M., Kavanagh T., Entian K.D., Rieger M., James R.,
 RA Pulgadenecch P., Hatzopoulos P., Obermaier B., Duestenhoft A.,
 RA Jones J., Palme K., Ansoorge W., Delseny M., Bancroft I., Mewes H.W.,
 RA Scheller C., Chalwatzis N.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z97336; CAB46059.1; -
 DR EMBL: AL161539; CAB78497.1; -
 DR InterPro: IPR003311; AUX_1AA.
 DR Pfam: PF02309; AUX_1AA; 1.
 DR SEQUENCE 234 AA; 25667 MW; 340A94E95B23F860 CRC64;
 SO

Query Match 55.0%; Score 44; DB 10; Length 234;
 Best Local Similarity 61.5%; Pred. No. 5.7;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 2 DODOMVVQVWMP 14
 Db 184 DGDWMLVGDVWMP 196

RESULT 7

ID 09XID4 PRELIMINARY; PRT; 619 AA.
 AC 09XID4;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE F23M19.4 protein.
 GN F23M19.4.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
 RA Lee J., Liu A., Li J., Kremetskaya I., Luros J., Gonzalez A.,
 RA Alrafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
 RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F23M19 sequence."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007454; AAC39615.1; -
 DR InterPro: IPR003311; AUX_1AA.
 DR InterPro: IPR003340; TF_B3.
 DR Pfam: PF02309; AUX_1AA; 1.
 DR Pfam: PF02362; B3; 1.
 DR SEQUENCE 619 AA; 6995 MW; 1DA4DD7BC43927E7 CRC64;
 SO

Query Match 55.0%; Score 44; DB 10; Length 619;
 Best Local Similarity 53.8%; Pred. No. 16;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 2 DODOMVVQVWMP 14
 Db 592 DEDKMLVGDVWMP 604

RESULT 8

ID 09LYN1 PRELIMINARY; PRT; 719 AA.
 AC 09LYN1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 77.3 kDa protein.
 GN F18021-60.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Benes V., Wurmbach E., Drzonek H., Ansoorge W., Mewes H.W., Rudd S.,
 RA Lemcke K., Mayer K.F.X., Queller F., Salanoubat M.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL163763; CAB87409.1; -
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 7.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PR000019; LEURICHRP.
 DR PRODOM: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00370; LRR; 6.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Hypothetical protein; Transferase.
 DR SEQUENCE 719 AA; 77329 MW; 3B91891A64E0DFD3 CRC64;
 SO

Query Match 55.0%; Score 44; DB 10; Length 719;
 Best Local Similarity 54.5%; Pred. No. 19;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 DQDMVVVQVWP 13
1 : : : 1 : 1 : 1 :
DB 100 QGQVIVITQLPW 110

RESULT 9

Q981L8 PRELIMINARY; PRT; 416 AA.
ID 0981L8
AC 0981L8
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Integrase/recombinase.
GN MUR3321.
RT Rhizobium loti (Mesorhizobium loti).
OC Plasmid pMIA.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_Taxid=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003015; BAB54691.1; -
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR002104; Phage_integrase.
DR Pfam: PF00589; Phage_integrase.1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN.1.
DR Plasmid: Complete proteome.
SQ SEQUENCE 416 AA; 46416 MW; 1DAB19A70297A655 CRC64;

Query Match 53.8%; Score 43; DB 16; Length 416;
Best Local Similarity 46.2%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDMVVVQVWP 14
1 : : : 1 : 1 : 1 :
DB 400 DADRLAAVALPW 412

RESULT 10

Q8UBJ3 PRELIMINARY; PRT; 418 AA.
ID 08UBJ3
AC 08UBJ3
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Intracellular PHB depolymerase.
GN DEPA OR ATU0015 OR AGR_C24.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_Taxid=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet M., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Senphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,

RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL: AE008976; AAL41046.1; -
DR EMBL: AE007944; AAK85840.1; -
KM Complete proteome.
SQ SEQUENCE 418 AA; 46630 MW; 666487C7A732B34 CRC64;

Query Match 53.8%; Score 43; DB 16; Length 418;
Best Local Similarity 87.5%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 VVVOVWP 14
1 : : : 1 : 1 : 1 :
DB 242 VVVOVWP 249

RESULT 11

Q81350 PRELIMINARY; PRT; 423 AA.
ID 081350
AC 081350
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE 5'-adenylsulfate reductase (EC 1.8.99.-).
OS Enteromorpha intestinalis (Hollow green seaweed).
OC Eukaryota; Viridiplantae; Chlorophyta; Ulvophyceae; Ulvales; Ulvaceae;
OC Enteromorpha.
OX NCBI_Taxid=3116;
RN [1]
RP SEQUENCE FROM N.A.
RA Gao Y., Leustek T.;
RT "Cloning of 5'-adenylsulfate (APS) reductase from the marine
RT macroalgae Enteromorpha intestinalis.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF069951; AAC26855.1; -
DR InterPro: IPR004508; APS_reduc.
DR InterPro: IPR002500; APS_reduc.
DR InterPro: IPR000063; Thioired.
DR Pfam: PF01507; PAPS_reduct.1.
DR Pfam: PF00085; Thioired.1.
DR PRINTS: PR00421; THIOREDOXIN.
DR TIGRFAMS: TIGR00424; APS_reduc.1.
KW Oxidoreductase.
SQ SEQUENCE 423 AA; 46808 MW; 808EE4F3BD561F2 CRC64;

Query Match 53.8%; Score 43; DB 10; Length 423;
Best Local Similarity 58.3%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 DQDMVVVQVWP 13
1 : : : 1 : 1 : 1 :
DB 330 DKDTWVLYAPW 341

RESULT 12

Q8S3B5 PRELIMINARY; PRT; 462 AA.
ID 08S3B5
AC 08S3B5;

DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Putative glucosyltransferase.
 OS Phaseolus lunatus (Lima bean) (Phaseolus limensis).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
 OX NCBI_TaxID=3884;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, KINGSTON;
 RA Shan X., Martin R.C., Mok D.W.S., Mok M.C.;
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF489877; AAM09517.1; -
 KW Transferase.
 SO SEQUENCE 462 AA; 51900 MW; 00C693CDF1E4646E CRC64;

Query Match 52.5%; Score 42; DB 10; Length 462;
 Best Local Similarity 53.8%; Pred. No. 28;
 Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 DODMNVVQVMP 14
 DB 10 DKNQVVVLPEP 22

RESULT 13

ID 039281 PRELIMINARY; PRT; 529 AA.
 AC 039281;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Counterpart of HSV-1 gene UL21 and VZV gene 38.
 GN 40.
 OS Equine herpesvirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicelloviruses.
 OX NCBI_TaxID=10331;
 RN [1]
 RP SEQUENCE OF 489-529 FROM N.A.
 RC STRAIN=NS80567;
 RX MEDLINE=90362066; PubMed=2167933;
 RA Nicolson L., Cullinane A.A., Onions D.E.;
 RT "The nucleotide sequence of an equine herpesvirus 4 gene homologue of the herpes simplex virus 1 glycoprotein H gene.";
 RL J. Gen. Virol. 71:1793-1800(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NS80567;
 RX MEDLINE=98264497; PubMed=9603335;
 RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
 RT "The DNA sequence of equine herpesvirus-4.";
 RL J. Gen. Virol. 79:1197-1203(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NS80567;
 RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
 RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF030027; AAC39558.1; -
 DR InterPro: IPR004936; Herpes_UL21.
 DR Pfam: PF03252; UL21; 1.
 SO SEQUENCE 529 AA; 58150 MW; E2B737B89D489413 CRC64;

Query Match 52.5%; Score 42; DB 12; Length 529;
 Best Local Similarity 45.5%; Pred. No. 32;
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 4 DODMNVVQVMP 14
 DB 283 DEILVEPMP 293

RESULT 14
 ID 09C8N7 PRELIMINARY; PRT; 600 AA.
 AC 09C8N7;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Auxin response factor, putative.
 GN F7P12.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
 RA Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Matzali A.,
 RA Miltner J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
 RL Nature 408:813-820(2000).
 DR EMBL: AC023913; AAC51894.1; -
 DR InterPro: IPR003311; AUX_1AA.
 DR InterPro: IPR003340; TF_B3.
 DR Pfam: PF02309; AUX_1AA; 1.
 DR Pfam: PF02362; B3; 1.
 SO SEQUENCE 600 AA; 67933 MW; C28A2B4E79073FFF CRC64;

Query Match 52.5%; Score 42; DB 10; Length 600;
 Best Local Similarity 53.8%; Pred. No. 37;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 DODMNVVQVMP 14
 DB 566 DDDKMLVGDPPMP 578

RESULT 15

ID 09LOE8 PRELIMINARY; PRT; 767 AA.
 AC 09LOE8;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE F1504.37.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Ecker J.R.;
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RA Ecker J.R.;
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
Khan S., Kim C., Altari H., Bel B., Chin C., Chiu J., Choi E.,
Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC007887; AAF79371.1; -
DR InterPro; IPR004160; EFTU_Cterm.
DR InterPro; IPR003340; TF_B3.
DR Pfam; PF02362; B3; 1
DR Pfam; PF03143; GTP_EFTU_D3; 1.
SQ SEQUENCE 767 AA; 85951 MW; FE29067F07D2E9F7 CRC64;

Query Match 52.5%; Score 42; DB 10; Length 767;
Best Local Similarity 46.2%; Pred. No. 48;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 DODQVVVVQVPP 14
:::|:|:| : |||
DB 621 EBDKMLVGEDPMP 633

Search completed: March 27, 2003, 10:06:01
Job time : 9.8536 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:04:25 ; search time 3.2439 seconds
(without alignments)
253.439 Million cell updates/sec

Title: US-09-675-208-3
Perfect score: 80
Sequence: 1 YDODQWVYQVWP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	55.0	830	10	US-09-772-134B-14
2	44	55.0	854	9	US-09-754-853A-1103
3	44	55.0	854	9	US-09-754-853A-1105
4	44	55.0	854	9	US-09-754-853A-1111
5	44	55.0	854	9	US-09-754-853A-1113
6	44	55.0	854	9	US-09-754-853A-1115
7	44	55.0	877	9	US-09-754-853A-1104
8	44	55.0	877	9	US-09-754-853A-1110
9	44	55.0	877	9	US-09-754-853A-1112
10	44	55.0	877	9	US-09-754-853A-1114
11	44	55.0	877	9	US-09-754-853A-1116
12	42	52.5	90	10	US-09-864-761-44178
13	40	50.0	46	9	US-09-798-889-50
14	40	50.0	854	9	US-09-754-853A-1098
15	40	50.0	854	9	US-09-754-853A-1101
16	40	50.0	854	9	US-09-754-853A-1107
17	40	50.0	854	9	US-09-754-853A-1109
18	40	50.0	877	9	US-09-754-853A-1097
19	40	50.0	877	9	US-09-754-853A-1100

20	40	50.0	877	9	US-09-754-853A-1106	Sequence 1106, Ap
21	40	50.0	877	9	US-09-754-853A-1108	Sequence 1108, Ap
22	38	47.5	404	10	US-09-908-805B-21	Sequence 21, Appl
23	38	47.5	526	10	US-09-855-154-2	Sequence 2, Appl1
24	38	47.5	574	10	US-10-025-380-1107	Sequence 1107, Ap
25	38	47.5	574	10	US-09-922-217-1107	Sequence 1107, Ap
26	38	47.5	722	9	US-10-029-495-5	Sequence 5, Appl1
27	37	46.2	33	10	US-09-281-717-50	Sequence 50, Appl
28	37	46.2	246	9	US-09-885-827-1	Sequence 1, Appl1
29	37	46.2	294	9	US-09-997-267-4	Sequence 4, Appl1
30	37	46.2	388	9	US-09-997-267-2	Sequence 2, Appl1
31	37	46.2	477	9	US-10-047-412A-10	Sequence 10, Appl
32	37	46.2	477	9	US-10-047-412A-29	Sequence 29, Appl
33	37	46.2	477	10	US-09-923-556-2	Sequence 2, Appl1
34	37	46.2	477	10	US-09-923-556-6	Sequence 6, Appl1
35	37	46.2	477	10	US-09-987-025-2	Sequence 2, Appl1
36	37	46.2	907	9	US-10-008-739A-2	Sequence 2, Appl1
37	37	46.2	978	10	US-09-893-817-4	Sequence 4, Appl1
38	37	46.2	978	10	US-09-893-817-8	Sequence 8, Appl1
39	36	45.0	191	9	US-09-950-933A-66	Sequence 66, Appl
40	36	45.0	229	9	US-09-975-719-265	Sequence 265, App
41	36	45.0	338	10	US-09-815-242-10944	Sequence 10944, A
42	36	45.0	400	9	US-09-943-688A-4	Sequence 4, Appl1
43	36	45.0	402	10	US-09-925-300-1320	Sequence 1320, Ap
44	36	45.0	404	9	US-09-943-689A-2	Sequence 2, Appl1
45	36	45.0	1152	9	US-09-738-626-3614	Sequence 3614, Ap

ALIGNMENTS

RESULT 1
US-09-772-134B-14
Sequence 14, Application US/09772134B
Patent No. US20020144310A1
GENERAL INFORMATION:
APPLICANT: Southern Illinois University
APPLICANT: Lightfoot, David
APPLICANT: Meksem, Khalid
TITLE OF INVENTION: ISOLATED POLYNUCLEOTIDES AND POLYPEPTIDES RELATING TO LOCI UND
TITLE OF INVENTION: RESISTANCE TO SOYBEAN CYST NEMATODE AND SOYBEAN SUDDEN DEATH
FILE REFERENCE: 1268/4/2
METHODS EMPLOYING SAME
CURRENT APPLICATION NUMBER: US/09/772.134B
CURRENT FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: 60/178,811
PRIOR FILING DATE: 2000-01-28
SOFTWARE: PatentIn version 3.0
SEQ ID NO 14
LENGTH: 830
TYPE: PRT
ORGANISM: soybean
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(830)
OTHER INFORMATION: xaa is any amino acid
US-09-772-134B-14
Query Match 55.0%; Score 44; DB 10; Length 830;
Best Local Similarity 54.5%; Pred. No. 10;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 3 QDQWVYQVWP 13
DB 24 QGQVIVIQLPW 34
RESULT 2
US-09-754-853A-1103
Sequence 1103, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:

```

; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1103
; LENGTH: 854
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_peeking_amplicon
US-09-754-853A-1103
```

```

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 854;
Best Local Similarity 54.5%; Pred. No. 10;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 3 ODOMVVVVVPM 13
1111111111
Db 115 OCGVIVIQLPW 125
```

```

RESULT 3
US-09-754-853A-1105
; Sequence 1105, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1105
; LENGTH: 854
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_toyosuzu_amplicon
US-09-754-853A-1105
```

```

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 854;
Best Local Similarity 54.5%; Pred. No. 10;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 3 ODOMVVVVVPM 13
1111111111
Db 115 OCGVIVIQLPW 125
```

```

RESULT 4
US-09-754-853A-1111
; Sequence 1111, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
```

```

; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1111
; LENGTH: 854
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_noir_amplicon
US-09-754-853A-1111
```

```

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 854;
Best Local Similarity 54.5%; Pred. No. 10;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 3 ODOMVVVVVPM 13
1111111111
Db 115 OCGVIVIQLPW 125
```

```

RESULT 5
US-09-754-853A-1113
; Sequence 1113, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1113
; LENGTH: 854
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_lee_amplicon
US-09-754-853A-1113
```

```

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 854;
Best Local Similarity 54.5%; Pred. No. 10;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 3 ODOMVVVVVPM 13
1111111111
Db 115 OCGVIVIQLPW 125
```

```

RESULT 6
US-09-754-853A-1115
; Sequence 1115, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
```



```
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1115
; LENGTH: 854
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_pl200499_amplicon
US-09-754-853A-1115

Query Match
Best Local Similarity 54.5%; Score 44; DB 9; Length 854;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ODOMVVVGVPM 13
DB 115 OGVIVIQLPW 125

RESULT 7
US-09-754-853A-1102
; Sequence 1102, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1102
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_pekling_amplicon
US-09-754-853A-1102

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 877;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ODOMVVVGVPM 13
DB 138 OGVIVIQLPW 148

RESULT 8
US-09-754-853A-1104
; Sequence 1104, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1104
; LENGTH: 877
; TYPE: PRT
```

```
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_toyosuzu_amplicon
US-09-754-853A-1104

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 877;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ODOMVVVGVPM 13
DB 138 OGVIVIQLPW 148

RESULT 9
US-09-754-853A-1110
; Sequence 1110, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1110
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_nolr_amplicon
US-09-754-853A-1110

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 877;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ODOMVVVGVPM 13
DB 138 OGVIVIQLPW 148

RESULT 10
US-09-754-853A-1112
; Sequence 1112, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1112
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_lee_amplicon
US-09-754-853A-1112
```

Query Match 55.0%; Score 44; DB 9; Length 877;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 QDQMVVVQVPM 13
I : : : : :
Db 138 QGQVIVIQLPW 148

RESULT 11
US-09-754-853A-1114
; Sequence 1114, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1114
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_pi200499_amplicon
US-09-754-853A-1114

Query Match 55.0%; Score 44; DB 9; Length 877;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 QDQMVVVQVPM 13
I : : : : :
Db 138 QGQVIVIQLPW 148

RESULT 12
US-09-864-761-44178
; Sequence 44178, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 44178
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005866.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EST_HUMAN HIT: AM962016.1, EVALUATE 2.00e-23
US-09-864-761-44178

Query Match 52.5%; Score 42; DB 10; Length 90;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 QAVVVQVPM 14
I : : : : :
Db 47 QLVVNFPM 56

RESULT 13
US-09-798-889-50
; Sequence 50, Application US/09798889
; Publication No. US20030004324A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 31 Human secreted proteins
; FILE REFERENCE: P2026P1
; CURRENT APPLICATION NUMBER: US/09/798,889
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/393,022
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,714
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,686
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,687
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,696
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Homo sapiens

```

FEATURE:
NAME/KEY: SITE
LOCATION: (46)
OTHER INFORMATION: Xaa equals stop translation
US-09-798-889-50

```

```

Query Match
Best Local Similarity 50.0%; Score 40; DB 9; Length 46;
Matches 6: Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 6 MYVVOVPM 14
:|:|:|:|:|
DB 18 LVVVSLLPM 26

```

```

RESULT 14
US-09-754-853A-1098
Sequence 1098, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:
APPLICANT: Hauge, Brian M.
APPLICANT: Parsons, Jeremy D.
APPLICANT: Parnell, Laurence D.
APPLICANT: Wang, Ming Li
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated with
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754,853A
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: US 60/174,880
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 1098
LENGTH: 854
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-1098

```

```

Query Match
Best Local Similarity 50.0%; Score 40; DB 9; Length 854;
Matches 5: Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 QMYYVOVPM 13
:|:|:|:|:|
DB 117 QVIVIQDLPW 125

```

```

RESULT 15
US-09-754-853A-1101
Sequence 1101, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:
APPLICANT: Hauge, Brian M.
APPLICANT: Parnell, Laurence D.
APPLICANT: Parsons, Jeremy D.
APPLICANT: Wang, Ming Li
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated with
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754,853A
PRIOR FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: US 60/174,880
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 1101
LENGTH: 854
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: rhg1_A3244_amplicon
US-09-754-853A-1101

```

```

Query Match
Best Local Similarity 50.0%; Score 40; DB 9; Length 854;
Matches 5: Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 QMYYVOVPM 13
:|:|:|:|:|
DB 117 QVIVIQDLPW 125

```

```

Search completed: March 27, 2003, 10:20:37
Job time: 4.2439 secs

```


GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:40 ; Search time 4.63415 Seconds
(without alignments)
120.634 Million cell updates/sec

Title: US-09-675-208-2
105
Sequence: 1 PVSVCFFVGFDPASPPSRH 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%
Maximum Match 10%
Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	19	4	US-08-846-234-2
2	105	100.0	784	4	US-08-846-234-5
3	42	40.0	22	4	US-08-905-223-460
4	41	39.0	212	4	US-08-861-774E-22
5	41	39.0	212	4	US-08-861-774E-34
6	40	38.1	291	4	US-09-105-697-4
7	40	38.1	291	4	US-09-105-697-5
8	40	38.1	291	4	US-09-105-697-6
9	40	38.1	834	1	US-07-977-434-8
10	40	38.1	834	1	US-07-977-434-10
11	40	38.1	834	1	US-08-073-384C-6
12	40	38.1	834	1	US-08-254-359A-6
13	40	38.1	834	1	US-08-384-490-31
14	40	38.1	834	1	US-08-483-043-6
15	40	38.1	834	1	US-08-459-383-31
16	40	38.1	834	1	US-08-458-819-8
17	40	38.1	834	1	US-08-458-819-10
18	40	38.1	834	1	US-08-481-238-6
19	40	38.1	834	2	US-08-471-066B-6
20	40	38.1	834	2	US-08-484-956-6
21	40	38.1	834	2	US-08-757-653-6
22	40	38.1	834	2	US-08-599-491-6
23	40	38.1	834	2	US-08-756-386-6
24	40	38.1	834	2	US-08-823-516-6
25	40	38.1	834	3	US-08-682-853A-6
26	40	38.1	834	3	US-08-759-038-6
27	40	38.1	834	3	US-08-758-314-6

28	40	38.1	834	4	US-09-350-309-6	Sequence 6, Appl1
29	40	38.1	834	4	US-08-520-946-6	Sequence 6, Appl1
30	40	38.1	834	5	PCT-US91-07035-8	Sequence 8, Appl1
31	40	38.1	834	5	PCT-US91-07035-10	Sequence 10, Appl1
32	40	38.1	834	5	PCT-US95-14418-2	Sequence 2, Appl1
33	40	38.1	834	5	PCT-US95-15327-2	Sequence 2, Appl1
34	39.5	37.6	435	2	US-08-338-530A-2	Sequence 2, Appl1
35	39.5	37.6	435	2	US-09-267-384-2	Sequence 2, Appl1
36	39	37.1	266	2	US-07-857-224B-39	Sequence 39, Appl1
37	39	37.1	481	4	US-08-843-659-4	Sequence 4, Appl1
38	39	37.1	560	5	PCT-US95-14418-5	Sequence 5, Appl1
39	39	37.1	560	5	PCT-US95-15327-5	Sequence 5, Appl1
40	39	37.1	597	5	PCT-US95-14418-4	Sequence 4, Appl1
41	39	37.1	597	5	PCT-US95-15327-4	Sequence 4, Appl1
42	39	37.1	833	1	US-08-073-384C-8	Sequence 8, Appl1
43	39	37.1	833	1	US-08-254-359A-8	Sequence 8, Appl1
44	39	37.1	833	1	US-08-483-043-8	Sequence 8, Appl1
45	39	37.1	833	1	US-08-481-238-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-846-234-2
; Sequence 2, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chleko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; US-08-846-234-2

Query Match 100.0%; Score 105; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVSVCFFVGFDPASPPSRH 19
DB 1 PVSVCFFVGFDPASPPSRH 19

```
RESULT 2
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-234-5

Query Match          100.0%; Score 105; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 1,2e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVSVCFCVGFDSSEPSR 19
Db 61 PVSVCFCVGFDSSEPSR 79

RESULT 3
US-08-905-223-460
; Sequence 460, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: win95
; SOFTWARE: word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
```

```
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 460:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig-peptide
; LOCATION: -17..-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 5.2
; OTHER INFORMATION: seq IMCLIGLKANASS/ET
US-08-905-223-460
```

```
Query Match          40.0%; Score 42; DB 4; Length 22;
Best Local Similarity 36.8%; Pred. No. 2,1;
Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
```

```
QY 1 PVSVCFCVGFDSSEPSR 19
Db 2 PVSIMCLIGLKANASSET 20
```

```
RESULT 4
US-08-861-774E-22
; Sequence 22, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Xanthoparmelia cumberlandia
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa-unknown amino acid
US-08-861-774E-22
```

```
Query Match          39.0%; Score 41; DB 4; Length 212;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 PVSVCFCVGFDSSEPD 16
Db 22 PEDVGCYIGACATDYD 37
```

```
RESULT 5
US-08-861-774E-34
; Sequence 34, Application US/08861774E
```


COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,697
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pcty Ph.D., Douglas A.
REGISTRATION NUMBER: 35321
REFERENCE/DOCKET NUMBER: 1043
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510)814-2974
TELEFAX: (510)814-2977
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 291 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-105-697-6

Query Match 38.1%; Score 40; DB 4; Length 291;
Best Local Similarity 61.5%; Pred. No. 75;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 7 FVGFDAEPDSRH 19
DB 64 FVFEAKAPSRH 76

RESULT 9
US-07-977-434-8
Sequence 8, Application US/07977434
Patent No. 5466591
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Adrarnson, Richard D.
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 7
SOFTWARE: wordperfect 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/977,434
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,490
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,466
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,213
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 523,394
FILING DATE: 15-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 143,441
FILING DATE: 12-JAN-1988
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 063,509
FILING DATE: 17-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,241
FILING DATE: 22-AUG-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 746,121
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO/PCT/US90/07641
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FILING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Luann Casert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2972
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-977-434-8

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2,4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
DB 64 FVFEAKAPSRH 76

RESULT 10
US-07-977-434-10
Sequence 10, Application US/07977434
Patent No. 5466591
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Adrarnson, Richard D.
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 7
SOFTWARE: wordperfect 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/977,434
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,490

FLING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,466
FLING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,213
FLING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 523,394
FLING DATE: 15-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 143,441
FLING DATE: 12-JAN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 063,509
FLING DATE: 17-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,241
FLING DATE: 22-AUG-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 746,121
FLING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US90/07641
FLING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FLING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FLING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FLING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FLING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Luann Caert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: Case No. 5465591 8753
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2972
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-977-434-10

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
DB 64 FVFDKAKPSFRH 76

RESULT 11
US-08-073-384C-6
Sequence 6, Application US/08073384C
Patent No. 5541311
GENERAL INFORMATION:
APPLICANT: Dahlberg, James E.
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL

STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,384C
FLING DATE: 04-JUN-1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FLING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-00613
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRADEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-073-384C-6

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
DB 64 FVFDKAKPSFRH 76

RESULT 12
US-08-254-359A-6
Sequence 6, Application US/08254359A
Patent No. 5614402
GENERAL INFORMATION:
APPLICANT: DAHLBERG, JAMES E.
APPLICANT: LYAMICHEV, VICTOR I.
APPLICANT: BROW, MARY ANN D.
TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,359A
FLING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FLING DATE: 06-JUN-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-254-359A-6

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAPSRH 76

RESULT 13

US-08-384-490-31
Sequence 31, Application US/08384490
Patent No. 5618711

GENERAL INFORMATION:

APPLICANT: Gelfand, David H.
APPLICANT: Scofield, Susanne
TITLE OF INVENTION: Recombinant Expression Vectors and
TITLE OF INVENTION: Purification Methods for Thermophilus DNA
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,490
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/148,133
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sias, Stacey R.
REGISTRATION NUMBER: 32,630
REFERENCE/DOCKET NUMBER: 8887
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2863
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-384-490-31

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAPSRH 76

RESULT 14

US-08-483-043-6
Sequence 6, Application US/08483043
Patent No. 5691142

GENERAL INFORMATION:

APPLICANT: Dahlberg, James E.
APPLICANT: Lyamchev, Victor I.
APPLICANT: Brow, Mary Ann D.
TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
TITLE OF INVENTION: POLYMERASE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERTOCK, MEDLEN & CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,043
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FILING DATE: 04-JUN-1993

GENERAL INFORMATION:

APPLICANT: Gelfand, David H.
APPLICANT: Scofield, Susanne
TITLE OF INVENTION: Recombinant Expression Vectors and
TITLE OF INVENTION: Purification Methods for Thermophilus DNA
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,043
FILING DATE:

GENERAL INFORMATION:

APPLICANT: Gelfand, David H.
APPLICANT: Scofield, Susanne
TITLE OF INVENTION: Recombinant Expression Vectors and
TITLE OF INVENTION: Purification Methods for Thermophilus DNA
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

OY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAPSRH 76

RESULT 15

US-08-459-383-31
Sequence 31, Application US/08459383
Patent No. 5741690

GENERAL INFORMATION:
APPLICANT: Gelfand, David H.

APPLICANT: Lawyer, Frances C.
APPLICANT: Scofield, Susanne
TITLE OF INVENTION: Recombinant Expression Vectors and
TITLE OF INVENTION: Purification Methods for Thermus Thermophilus DNA
TITLE OF INVENTION: Polymerase
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,383
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/384,490
FILING DATE:
APPLICATION NUMBER: US/08/148,133
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sias, Stacey R.
REGISTRATION NUMBER: 32,630
REFERENCE/DOCKET NUMBER: 8867
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2863
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-459-383-31

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0;
Gaps 0;

QY 7 FVGFDAEPDSRH 19
DB 64 FVFDKAPSPFRH 76

Search completed: March 27, 2003, 10:07:34
Job time : 5.63415 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 13.439 Seconds
(without alignments)
188.389 Million cell updates/sec

Title: US-09-675-208-2

Perfect score: 105

Sequence: 1 PVSVCFFVGFDPASEPDSRH 19

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

```
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	19	AAW53568	Cucurbit raffinose
2	105	100.0	19	AAV17423	Cucurbit raffinose
3	105	100.0	784	AAW53570	Cucurbit raffinose
4	105	100.0	784	AAV17417	Cucurbit raffinose
5	71	67.6	781	AAW57887	Soybean protein: S
6	71	67.6	781	AAW57887	Soybean raffinose
7	71	67.6	781	AAV30143	Amino acid sequenc
8	71	67.6	781	AAW54400	Soybean raffinose
9	63	60.0	799	AAW57886	Broad bean raffin
10	63	60.0	799	AAV30142	Amino acid sequenc

11	62	59.0	758	21	AAV70978
12	61	58.1	783	20	AAV32073
13	47.5	45.2	770	21	AAV70976
14	47	44.8	175	21	AAW18865
15	47	44.8	253	21	AAW18864
16	47	44.8	295	21	AAW18863
17	46.5	44.3	777	20	AAV32074
18	46.5	44.3	783	23	ABW93664
19	45	42.9	763	21	AAV70977
20	44	41.9	112	18	AAW08433
21	44	41.9	112	22	AAW63148
22	44	41.9	428	22	ABW60476
23	43	41.0	86	23	ABP31035
24	43	41.0	313	22	AAW50675
25	43	41.0	441	20	AAV40501
26	43	41.0	447	21	AAW15496
27	42.5	40.5	841	21	AAV70981
28	42	40.0	22	20	AAV13176
29	42	40.0	257	17	AAW94460
30	42	40.0	382	22	AAW93040
31	42	40.0	382	22	AAW76725
32	41	39.0	212	20	AAW82677
33	41	39.0	212	20	AAW82678
34	41	39.0	212	20	AAW82685
35	41	39.0	823	22	AAW19520
36	40	38.1	48	22	AAO10988
37	40	38.1	52	22	ABG01020
38	40	38.1	52	22	ABG01021
39	40	38.1	78	22	AAU47171
40	40	38.1	129	22	AAW00853
41	40	38.1	144	20	AAV59807
42	40	38.1	145	20	AAV48481
43	40	38.1	146	22	ABG20734
44	40	38.1	232	22	AAW82256
45	40	38.1	283	23	ABW57205

ALIGNMENTS

RESULT 1
AAW53568
ID AAW53568 standard; peptide; 19 AA.
XX
AC AAW53568;
XX
06-JUL-1998 (first entry)
XX
DE Cucurbit raffinose synthase residues 61 to 79.
XX
XX Cucurbit raffinose synthase; sucrose; galactinol.
XX
OS Cucumis sativus.
XX
PN JPI0084973-A.
XX
PD 07-APR-1998.
XX
PE 28-APR-1997; 97JP-011124.
XX
PR 26-JUL-1996; 96JP-0198079.
PR 26-APR-1996; 96JP-0107682.
XX
PA (AJIN) AJINOMOTO KK.
XX
DR WPI; 1998-264858/24.
XX
PT Raffinose synthase gene - useful for preparation of raffinose in
PT transformed plant
PS Example 2; Page 17; 26pp; Japanese.
XX
XX The present sequence is a cucurbit raffinose synthase fragment.

CC Raffinose synthase forms raffinose from sucrose and galactinol, has
 CC an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees
 CC C, has a molecular weight of 75 to 95 kDa by gel filtration or 90
 CC to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is
 CC inhibited by Iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 19 AA;

Query Match 100.0%; Score 105; DB 19; Length 19;
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCFCVGFDASEPDSSRH 19
 |||||
 DB 1 PVSVCFCVGFDASEPDSSRH 19

RESULT 2

AAV17423
 ID AAV17423 standard; peptide; 19 AA.

XX AAV17423;

XX 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase peptide SEQ ID NO:2.

XX Raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

XX JP1123080-A.

XX 11-MAY-1999.

XX 24-OCT-1997; 97JP-0292969.

XX 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1999-340516/29.

PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

PS Example 2; Page 21; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents a raffinose
 CC synthase peptide from cucumber.

XX Sequence 19 AA;

Query Match 100.0%; Score 105; DB 20; Length 19;
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCFCVGFDASEPDSSRH 19
 |||||
 DB 1 PVSVCFCVGFDASEPDSSRH 19

RESULT 3

AAW53570
 ID AAW53570 standard; Protein; 784 AA.

XX AAW53570;

XX 06-JUL-1998 (first entry)

XX Cucumber raffinose synthase.

XX Cucumber: raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

XX JP10084973-A.

XX 07-APR-1998.

XX 28-APR-1997; 97JP-0111124.

XX 26-JUL-1996; 96JP-0198079.

XX 26-APR-1996; 96JP-0107682.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1998-264858/24.

XX N-PSDB; AAV22250.

PT Raffinose synthase gene - useful for preparation of raffinose in
 PT transformed plant

PS Claim 3; Pages 17-20; 26pp; Japanese.

CC The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC Iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA;

Query Match 100.0%; Score 105; DB 19; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCFCVGFDASEPDSSRH 19
 |||||
 DB 61 PVSVCFCVGFDASEPDSSRH 79

RESULT 4

AAV17417
 ID AAV17417 standard; Protein; 784 AA.

XX AAV17417;

XX 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase.

XX Raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

XX JP1123080-A.

XX 11-MAY-1999.

XX 24-OCT-1997; 97JP-0292969.

XX 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1999-340516/29.

XX N-PSDB; AAX61238.

PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

XX Claim 2; Page 25-27; 37pp; Japanese.
 PS
 XX
 CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents raffinose
 CC synthase from cucumber.
 XX
 SQ Sequence 784 AA;

Query Match 100.0%; Score 105; DB 20; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCFGVGFDASEPDSSRH 19
 ||||||||||||||||
 DB 61 PVSVCFGVGFDASEPDSSRH 79

RESULT 5

AAB98659
 ID AAB98659 standard; protein; 780 AA.

XX
 AC AAB98659;

DT 17-AUG-2001 (first entry)

DE Soybean protein; SEQ ID 1.

KW Mutant; muteln; raffinose synthase; raffinose oligosaccharide reduction;
 plant; soybean.

XX
 OS Glycine max.

PN JP2001078783-A.

XX
 PD 27-MAR-2001.

PF 03-JUL-2000; 2000JP-0200571.

XX
 PR 09-JUL-1999; 99JP-0196036.

XX
 PA (SUMO) SUMITOMO CHEM CO LTD.

DR WPI: 2001-313373/33.

XX
 DR N-PSDB: AAH27438.

PT Novel mutant protein of raffinose synthase is useful for reducing the
 raffinose oligosaccharide content in a plant body -

XX
 PS Disclosure: Page 18-20; 30pp; Japanese.

XX
 CC The present invention relates to a mutant protein of raffinose synthase
 CC in which at least one aromatic amino acid present at the position of
 CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
 CC mutant protein can be used for reducing the raffinose oligosaccharide
 CC content in a plant body. The present protein from soybean, was used in
 CC the present invention.
 XX

SQ Sequence 780 AA;

Query Match 67.6%; Score 71; DB 22; Length 780;
 Best Local Similarity 81.2%; Pred. No. 0.0044;
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCFCVGFDASEPDSSRH 19
 ||||||| | | | | |
 DB 68 VGCFCVGFDADEPRSRH 83

RESULT 6

AAW57887
 ID AAW57887 standard; Protein; 781 AA.

XX
 AC AAW57887;

DT 23-SEP-1998 (first entry)

XX
 DE Soybean raffinose synthetase.

XX
 KW Raffinose synthetase; metabolism modification; food additive;
 gastrointestinal flora; soybean.

XX
 OS Glycine max.

PN EP849359-A2.

XX
 PD 24-JUN-1998.

PF 18-DEC-1997; 97EP-0122417.

XX
 PR 18-DEC-1996; 96JP-0338673.

XX
 PA (SUMO) SUMITOMO CHEM CO LTD.

XX
 PI Oeda K, Wantanabe E;

DR WPI: 1998-324670/29.

XX
 DR N-PSDB: AAW40801.

PT New nucleic acid molecule encoding plant raffinose synthetase -
 capable of producing raffinose, used as food additives with
 beneficial effects on gastrointestinal flora

XX
 PS Claim 1; Page 31-34; 44pp; English.

XX
 CC This sequence represents the soybean raffinose synthetase of the
 CC invention. The raffinose synthetase is capable of producing raffinose by
 CC combining a D-galactosyl group through an alpha (1-6) bond with a
 CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
 CC residue in a sucrose molecule. The DNA can be used to modify metabolism
 CC of a host organism by introducing into the host organism or its cell so
 CC that the content of the raffinose family oligosaccharides in the host
 CC organism or cell is changed. Raffinose oligosaccharides are useful as
 CC food additives with beneficial effects on the gastrointestinal flora.

XX
 SQ Sequence 781 AA;

Query Match 67.6%; Score 71; DB 19; Length 781;
 Best Local Similarity 81.2%; Pred. No. 0.0044;
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCFCVGFDASEPDSSRH 19
 ||||||| | | | | |
 DB 68 VGCFCVGFDADEPRSRH 83

RESULT 7

AAV30143
 ID AAV30143 standard; Protein; 781 AA.

XX
 AC AAV30143;

DT 26-OCT-1999 (first entry)

XX
 DE Amino acid sequence of a raffinose synthase protein.

XX
 KW Raffinose synthase; plant; sucrose; raffinose.

XX
 OS Glycine max.

PN JPI1215984-A.

XX
 PD 10-AUG-1999.

```

XX 12-DEC-1997; 97JP-0342899.
XX
PF 28-NOV-1997; 97JP-0329006.
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR WPI: 1999-51112/43.
DR N-PSDB; AAI10002.
XX
PT New raffinose synthase gene - is prepared from a plant material
XX
PS Claim 8; Page 25-27; 40pp; Japanese.
XX
CC The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6)-D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
SQ Sequence 781 AA:

Query Match 67.6%; Score 71; DB 20; Length 781;
Best Local Similarity 81.2%; Pred. No. 0.0044;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCVFGEFADSEPDPSRH 19
DB 68 VGCVFGEFADSEPDPSRH 83

RESULT 8
AAB49400
ID AAB49400 standard; Protein; 781 AA.
XX
AC AAB49400;
XX
DT 07-MAR-2001 (first entry)
XX
DE Soybean raffinose synthase.
XX
KW Plant promoter; transgenic plant; desired property.
XX
OS Glycine max.
XX
PN EP1048733-A2.
XX
PD 02-NOV-2000.
XX
PF 27-APR-2000; 2000EP-0108962.
XX
PR 30-APR-1999; 99JP-0124527.
PR 01-SEP-1999; 99JP-0247211.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Ishige F, Watanabe E, Oeda K;
XX
DR WPI: 2001-104537/12.
DR N-PSDB; AAC89523.
XX
PT New soybean plant promoters useful for generating transgenic plants
PT with desired properties -
XX
PS Example 6; Page 24-27; 36pp; English.
XX
CC The present invention provides novel plant promoters which can be used in
CC the production of transgenic plants which express genes with desired
CC properties.
XX
SQ Sequence 781 AA:

Query Match 67.6%; Score 71; DB 22; Length 781;

```

```

Best Local Similarity 81.2%; Pred. No. 0.0044;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCVFGEFADSEPDPSRH 19
DB 68 VGCVFGEFADSEPDPSRH 83

RESULT 9
AAW57886
ID AAW57886 standard; Protein; 799 AA.
XX
AC AAW57886;
XX
DT 23-SEP-1998 (first entry)
XX
DE Broad bean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; broad bean.
XX
OS Vicia faba.
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.
XX
PF 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Watanabe E;
XX
DR WPI: 1998-324670/29.
DR N-PSDB; AAV40800.
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose; used as food additives with
PT beneficial effects on gastrointestinal flora
XX
PS Claim 1; Page 26-29; 44pp; English.
XX
CC This sequence represents the broad bean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
SQ Sequence 799 AA:

Query Match 60.0%; Score 63; DB 19; Length 799;
Best Local Similarity 66.7%; Pred. No. 0.092;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 GCFVGFADSEPDPSRH 19
DB 87 GCFVGFADSEPDPSRH 101

RESULT 10
AAV30142
ID AAV30142 standard; Protein; 799 AA.
XX
AC AAV30142;
XX
DT 26-OCT-1999 (first entry)
XX

```


DE Amino acid sequence of a raffinose synthase protein.
 XX
 XX Raffinose synthase; plant; broad bean; sucrose; raffinose.
 KW
 OS Vicia faba.
 XX
 PN JP11215984-A.
 PD 10-AUG-1999.
 XX
 PE 12-DEC-1997; 97JP-0342899.
 XX
 PR 28-NOV-1997; 97JP-0329006.
 PR 18-DEC-1996; 96JP-0338673.
 XX
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI: 1999-511112/43.
 DR N-PSDB: AA210001.
 XX
 PT New raffinose synthase gene - is prepared from a plant material
 PS
 XX Claim 5; Page 19-21; 40pp; Japanese.
 CC The present sequence represents a raffinose synthase protein. The
 CC sequence is isolated from plant material of broad beans. The
 CC protein forms raffinose by complexing alpha(1 to 6)- D-galactosyl
 CC hydroxyl group of the 6C of D-glucose residue in sucrose molecules.
 XX
 SQ Sequence 799 AA:
 OY 5 GCFVGFDPSEPSRH 19
 DB 87 GCFVGFDPSEPSRH 101
 Query Match 60.0%; Score 63; DB 20; Length 799;
 Best Local Similarity 66.7%; Pred. No. 0.092;
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

RESULT 11
 AAY70978
 ID AAY70978 standard; Protein: 758 AA.
 AC AAY70978;
 XX
 DT 09-AUG-2000 (first entry)
 XX
 DE Soybean raffinose synthase from clone sfl1.pk125.d4.
 XX
 KW Soybean: raffinose synthase; raffinose saccharide;
 KW clone sfl1.pk125.d4; nutritional; soy protein.
 XX
 OS Glycine max.
 XX
 PN WO200024915-A2.
 PD 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99WO-US24923.
 XX
 PR 23-OCT-1998; 98US-0105451.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX
 PI Allen SM, Hiltz WD;
 DR WPI: 2000-350754/30.
 DR N-PSDB: AAD00335.
 XX
 PT Nucleic acids and encoded proteins involved in the biosynthesis of
 PT raffinose, useful for producing soybean seeds with a reduced raffinose
 PT content and therefore improved nutritional quality -

XX
 PS Claim 2; Page 47-49; 58pp; English.
 XX
 CC The present sequence is a raffinose synthase from
 CC clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
 CC library sfl1. Raffinose synthase is involved in the biosynthesis
 CC of raffinose and higher homologues in the raffinose saccharide family
 CC from sucrose. The present sequence is useful for reducing the raffinose
 CC saccharide content of soybean seeds which improves the nutritional
 CC quality of the soy protein products derived from them.
 XX
 SQ Sequence 758 AA:
 OY 3 SVGCFVGFDPSEPSRH 19
 DB 54 TTGCFGFGHATSPKSRH 70
 Query Match 59.0%; Score 62; DB 21; Length 758;
 Best Local Similarity 58.8%; Pred. No. 0.13;
 Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

RESULT 12
 AAY32073
 ID AAY32073 standard; Protein: 783 AA.
 AC AAY32073;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Sugarbeet raffinose synthase.
 XX
 KW Raffinose synthase; sugarbeet; transgenic plant.
 XX
 OS Beta vulgaris.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 227
 FT Misc-difference 227 /note= "encoded by CCW"
 FT Misc-difference 348 /note= "encoded by CCR"
 XX
 PN EP953643-A2.
 PD 03-NOV-1999.
 XX
 PF 27-APR-1999; 99EP-0107430.
 XX
 PR 30-APR-1998; 98JP-0120550.
 PR 30-APR-1998; 98JP-0120551.
 PR 04-DEC-1998; 98JP-0345590.
 PR 10-DEC-1998; 98JP-0351246.
 XX
 PA (SUMO) SUMITOMO CHEM CO LTD.
 XX
 PI Watanabe E, Oeda K;
 DR WPI: 1999-593144/51.
 DR N-PSDB: AA220208.
 XX
 PT New sense and antisense genes, useful for altering the level of
 PT raffinose in food plants -
 PS
 PS Claim 25; Page 22-24; 55pp; English.
 CC This sequence represents sugarbeet raffinose synthase, a protein
 CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
 CC hydroxy group attached to the carbon atom at the 6-position of the
 CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
 CC (see AA220208) encoding the enzyme was isolated from sugarbeet cv.
 CC hanning leaf cDNA by PCR. Probes or primers generated from plant
 CC raffinose synthase genes (see AA220207-10) may be used to obtain
 CC other raffinose synthase genes by labeled detection or amplification

CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
XX

XX Sequence 783 AA;

Query Match 58.1%; Score 61; DB:20; Length 783;

Best Local Similarity 73.3%; Pred No. 0.19;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 GCFVGFDAEPDSRH 19

DB 60 GLFVGFDAPFKARH 74

RESULT 13

AAV70976

ID AAV70976 standard; Protein: 770 AA.

XX AAV70976;

DT 09-AUG-2000 (first entry)

XX Rice raffinose synthase from clone rls24.pk0017.g10.

XX Rice; raffinose synthase; raffinose saccharide; soybean;

KW clone rls24.pk0017.g10; nutritional; soy protein.

XX Oryza sativa.

XX Key Location/Qualifiers

FT MISC-difference 100..101

FT MISC-difference 101..110

FT MISC-difference 101..110

FT /label="Unknown

FT /note="Encoded by NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN"

XX WO200024915-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24923.

XX 23-OCT-1998; 98US-0105451.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Allen SM, Hiltz WD;

XX WPI: 2000-350754/30.

XX N-PSDB: AAD00333.

XX Nucleic acids and encoded proteins involved in the biosynthesis of

XX raffinose, useful for producing soybean seeds with a reduced raffinose

XX content and therefore improved nutritional quality -

XX Claim 2; Page 40-42; 58pp; English.

XX The present sequence is a raffinose synthase from

XX clone rls24.pk0017.g10 isolated from a rice infected leaf cDNA

XX library rls24. Raffinose synthase is involved in the biosynthesis

XX of raffinose and higher homologues in the raffinose saccharide family

XX from sucrose. The present sequence is useful for reducing the raffinose

XX saccharide content of soybean seeds which improves the nutritional

XX quality of the soy protein products derived from them.

Best Local Similarity 57.9%; Pred. No. 30;

Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

OY 1 PVSVCVGFDAEPDSRH 19

DB 39 PVD-GVFLGDFAPKARH 56

RESULT 14

AAAG16865

ID AAAG16865 standard; Protein: 175 AA.

XX AAAG16865;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 17671.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130440.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 07-MAY-1999; 99US-0132486.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

1;

1;

1;

56

14

AAAG16865

ID

XX

DT

XX

XX

KW

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 44.8%; Score 47; DB 21; Length 175;
Best Local Similarity 64.3%; Pred. No. 8;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCFVGFDSRPSDR 18
DB 90 GEFVGFDSRPSDR 103

RESULT 15
AAG16864
ID AAG16864 standard; Protein; 253 AA.
XX
AC AAG16864;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17670.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
OS
PN EPI033405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825;
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0132856.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158237.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161932.
PR	29-OCT-1999;	99US-0161933.
PR	29-OCT-1999;	99US-0162142.

Query Match	44.8%	Score 47	DB 21	Length 253
Best Local Similarity	64.3%	Pred. No. 12		
Matches	9	Conservative	2	Mismatches 3
				Indels 0
				Gaps 0
OY	5	GCFVGFDA	SEPSR	18
db	168	GEFLGFDENE	PTSR	181

Search completed: March 27, 2003, 10:03:43
Job time : 15.439 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 5.44512 Seconds
(without alignments)
335.448 Million cell updates/sec

Title: US-09-675-208-2

Perfect score: 105
Sequence: 1 PVSVCFCVGFDPASEPDSRH 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	46.7	836	2 AF3233	conserved hypotet
2	47	44.8	295	2 B66320	3-phosphoserine ph
3	47	44.8	295	2 T51362	phosphoserine phos
4	47	44.8	336	2 A97544	nifH3/asm1 family
5	47	44.8	336	2 AB2763	conserved hypotet
6	46	43.8	483	2 AH3445	GTP-binding protei
7	44	41.9	112	2 S68584	hydrophobin A - cu
8	44	41.9	301	2 S73023	hypothetical prote
9	44	41.9	428	2 A43741	terminus protein -
10	44	41.9	498	2 T48269	hypothetical prote
11	44	41.9	650	1 JCI450	fibroblast growth
12	43	41.0	313	2 T33185	hypothetical prote
13	43	41.0	441	2 F66185	hypothetical prote
14	43	41.0	447	2 T02547	hypothetical prote
15	43	41.0	464	2 S58898	kynureninase (EC 3
16	43	41.0	464	2 T48675	kynureninase (EC 3
17	43	41.0	478	2 T15516	hypothetical prote
18	42.5	40.5	546	2 A69890	hypothetical prote
19	42	40.0	72	2 T46623	hypothetical prote
20	42	40.0	169	2 G72559	hypothetical prote
21	42	40.0	240	2 E83896	hypothetical prote
22	42	40.0	311	2 T32776	hypothetical prote
23	42	40.0	314	2 T27383	hypothetical prote
24	42	40.0	364	2 AD2302	hypothetical prote
25	42	40.0	511	2 T23459	hypothetical prote
26	42	40.0	653	2 B87698	transketolase I [l
27	41.5	39.5	219	2 T42605	envelope protein -
28	41	39.0	125	2 E81192	hypothetical prote
29	41	39.0	314	2 T25842	hypothetical prote

30	41	39.0	511	2 G87609	L-aspartate oxidase
31	41	39.0	552	2 T16345	hypothetical prote
32	41	39.0	5232	2 A45086	HC-toxin synthetase
33	40.5	38.6	498	2 A83635	conserved hypotet
34	40	38.1	80	2 G86417	unknown protein, 6
35	40	38.1	104	2 S36510	E7 protein - human
36	40	38.1	260	2 G86228	hypothetical prote
37	40	38.1	282	2 A36875	plasmalemmal volta
38	40	38.1	362	2 A84187	hypothetical prote
39	40	38.1	447	2 T50705	hypothetical prote
40	40	38.1	465	2 G02652	gamma-aminobutyrate
41	40	38.1	520	2 AD2383	kynureninase (EC 3
42	40	38.1	609	2 E82423	Na+/H+-exchanging
43	40	38.1	615	1 ABCHS	hypothetical prote
44	40	38.1	650	2 A54976	serum albumin prec
45	40	38.1	755	2 B41836	peroxisomal matrix
					amine oxidase (fla

ALIGNMENTS

RESULT 1
AF3233
conserved hypothetical protein Atu6048 [Imported] - Agrobacterium tumefaciens (strain C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AF3233
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Genthner, D.; Kutayavin, T.; Levy, R.; Li, M.; McGill
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF3233
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-836 <KUR>
A:Cross-references: GB:AE008690; PIDN:AAL46284.1; PID:q17744066; GSPDB:GN00189
A:Experimental source: strain C58 (dupont)
C:Genetics:
A:Gene: Atu6048
A:Genome: plasmid
C:Superfamily: Lactococcus lactis hypothetical protein PL08712.1

Query Match 46.7%; Score 49; DB 2; Length 836;
Best Local Similarity 41.2%; Pred. No. 6.5;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 3 SVSCFCVGFDPASEPDSRH 19
DB 687 NLGATVGFDPAREPAVH 703

RESULT 2
B66320
3-phosphoserine phosphatase [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B66320
R:Rheolids, A.; Ecker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dekar,
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712
A:Accession: B66320

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-295 <STO>
 A:Cross-references: GB:AE005172; NID:g97955592; PIDN:AAF98410.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 44.8%; Score 47; DB 2; Length 295;
 Best Local Similarity 64.3%; Pred. No. 4.8;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCFVGFDAEPDSR 18
 DB 210 GEFLEGFDENEPTSR 223

RESULT 3
 T51362
 phosphoserine phosphatase (EC 3.1.3.3) precursor, chloroplast [validated] - Arabidopsis
 M:Alternate names: 3-phosphoserine phosphatase
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
 C:Accession: T51362
 R:HO, C.; Noji, M.; Saito, K.
 J. Biol. Chem. 274, 11007-11012, 1999
 A:Title: Plasticidic pathway of serine biosynthesis. Molecular cloning and expression of
 A:Reference number: Z25385
 A:Accession: T51362
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-295 <HOC>
 A:Cross-references: EMBL:AB018408; PIDN:BAA33806.1
 C:Genetics:
 A:Gene: PSP
 A:Map position: 1
 A:Genome: nuclear
 C:Function:
 A:Description: (EC 3.1.3.3) [validated, MUID:99214175]
 A:Pathway: serine biosynthesis (plasticidic pathway)
 C:Keywords: chloroplast; phosphoric monoester hydrolase

Query Match 44.8%; Score 47; DB 2; Length 295;
 Best Local Similarity 64.3%; Pred. No. 4.8;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCFVGFDAEPDSR 18
 DB 210 GEFLEGFDENEPTSR 223

RESULT 4
 A97544
 nif3/smm1 family protein WC0379 [imported] - Agrobacterium tumefaciens (strain C58, Cer
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: A97544
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollm, C.; Allinger, M.; Doughty, D.; Scott, C.; Iappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; PMID:11743194
 A:Accession: A97544
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-336 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK87306.1; PID:g15156601; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_2794
 A:Map position: circular chromosome
 C:Superfamily: Haemophilus Influenzae conserved hypothetical protein HI0634

Query Match 44.8%; Score 47; DB 2; Length 336;
 Best Local Similarity 47.1%; Pred. No. 5.5;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCVGFDAEPDS 17
 DB 135 PVTWKCRIGVDDDEPT 151

RESULT 5
 AB2763
 conserved hypothetical protein Atu1515 [imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
 C:Accession: AB2763
 R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavh, T.; Levy, R.; Li, M.; MCI
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AB2763
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-336 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAL42520.1; PID:g17739940; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu1515
 A:Map position: circular chromosome
 C:Superfamily: Haemophilus Influenzae conserved hypothetical protein HI0634

Query Match 44.8%; Score 47; DB 2; Length 336;
 Best Local Similarity 47.1%; Pred. No. 5.5;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCVGFDAEPDS 17
 DB 135 PVTWKCRIGVDDDEPT 151

RESULT 6
 AH3445
 GTP-binding protein [imported] - Brucella melitensis (strain 16M)
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AH3445
 R:DeIvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
 .; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Let
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AH3445
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-483 <KUR>
 A:Cross-references: GB:AE008917; PIDN:AAL52731.1; PID:g17983561; GSPDB:GN00190
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BME11550
 A:Map position: I
 C:Superfamily: Mycobacterium leprae probable GTP-binding protein: translation elongat

Query Match 43.8%; Score 46; DB 2; Length 483;
 Best Local Similarity 47.1%; Pred. No. 12;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PVSVCVGFDAEPDS 17
 DB 184 PAATGALVDDDIEDPDA 200

RESULT 7

S68584
hydrophobin A - cultivated mushroom
C:Species: Agaricus bisporus (cultivated mushroom)
C:Date: 06-Dec-1996 #sequence_revision 07-Feb-1997 #text_change 21-Jul-2000
C:Accession: S68584; S68586; S58342
R:De Groot, P.W.J.; Schaap, P.J.; Sonnenberg, A.S.M.; Visser, J.; van Gilsen, L.J.L.D
J. Mol. Biol. 257, 1008-1018, 1996
A:Title: The Agaricus bisporus hypha gene encodes a hydrophobin and specifically accumulates
A:Reference number: S68584; MUID:96192085; PMID:8632464
A:Accession: S68584
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-112 <DEC>
A:Cross-references: EMBL:X90818; NID:g1235753; PIDN:CAA62331.1; PID:g1235754
A:Accession: S68586
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-112 <DEM>
A:Cross-references: EMBL:X89242; NID:g944820; PID:g944821
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
C:Genetics:
A:Gene: hypha
A:Introns: 18/3; 40/1; 96/3
C:Superfamily: 1G2 protein

Query Match 41.9%; Score 44; DB 2; Length 112;
Best Local Similarity 46.2%; Pred. No. 5.5;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 PVSVCVGFVGDAS 13
I::I I::I I::I
DB 60 PNLGAFLEFDCT 72

RESULT 8
S73023
hypothetical protein L518_F3_81 - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C:Accession: S73023
R:Smith, D.R.; Robinson, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid L518.
A:Reference number: S72591
A:Accession: S73023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <SMI>
A:Cross-references: EMBL:U00023; NID:g467194; PIDN:AAA17366.1; PID:g467209
C:Superfamily: Mycobacterium leprae hypothetical protein L518_F3_81

Query Match 41.9%; Score 44; DB 2; Length 301;
Best Local Similarity 57.1%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 PVSVCVGFVGDASE 14
I::I I::I I::I
DB 200 PARCGCFVGVDFTE 213

RESULT 9
AA3741
terminus protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 04-Mar-2000
C:Accession: AA3741
R:Balderelli, R.M.; Mahoney, P.A.; Salas, F.; Gustavson, E.; Boyer, P.D.; Chang, M.F.; F
Dev. Biol. 125, 85-95, 1988
A:Title: Transcripts of the Drosophila blastoderm-specific locus, terminus, are concentr
A:Reference number: AA3741; MUID:88055884; PMID:3334721
A:Accession: AA3741
A:Status: preliminary
A:Molecule type: mRNA

A:Residues: 1-428 <BAL>
A:Cross-references: EMBL:M19140; NID:g158535; PID:g158536
C:Genetics:
A:Gene: FlyBase:term
A:Cross-references: FlyBase:Fbgn0003683
C:Superfamily: Drosophila terminus protein
C:Keywords: DNA binding

Query Match 41.9%; Score 44; DB 2; Length 428;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 6 CFVGFADSEPDSSRH 19
I::I I::I I::I
DB 37 CYSIDADPEPSQH 50

RESULT 10
T48269
hypothetical protein T22P11.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T48269
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, April 2000
A:Reference number: T24490
A:Accession: T48269
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-498 <BEV>
A:Cross-references: EMBL:AL162971
A:Experimental source: cultivar Columbia; BAC clone T22P11
C:Genetics:
A:Map position: 5
A:Note: T22P11.70
C:Superfamily: Arabidopsis thaliana hypothetical protein T13M22.7

Query Match 41.9%; Score 44; DB 2; Length 498;
Best Local Similarity 70.0%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 5 GCFVGFADSE 14
I::I I::I I::I
DB 132 GCFTGFDKSD 141

RESULT 11
JC1450
fibroblast growth factor receptor 4 - rat
N:Contains: protein-tyrosine kinase (EC 2.7.1.112)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 01-Dec-2000
C:Accession: JC1450; PT0191
R:Horlick, R.A.; Stack, S.L.; Cooke, G.M.
Gene 120, 291-295, 1992
A:Title: Cloning, expression and tissue distribution of the gene encoding rat fibrobl
A:Reference number: JC1450; MUID:93013049; PMID:1398143
A:Accession: JC1450
A:Molecule type: mRNA
A:Residues: 1-650 <HOR>
A:Cross-references: GB:M91599; NID:g204137; PIDN:AAA41157.1; PID:g204138
R:Ial, C.; Lemke, G.
Neuron 6, 691-704, 1991
A:Title: An extended family of protein-tyrosine kinase genes differentially expressed
A:Reference number: PT0185; MUID:91222560; PMID:2023425
A:Accession: PT0191
A:Molecule type: mRNA
A:Residues: 465-518 <LAI>
A:Experimental source: sciatic nerve
C:Genetics:
A:Gene: FGFR4; tyro-9
A:Function:
A:Description: receptor mediating effects of fibroblast growth factor

A>Note: expressed in normal lung; expressed in some carcinomas
 C:Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; protb
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; growth factor receptor;
 F:11-72/Domain: immunoglobulin homology <IM1>
 F:110-181/Domain: immunoglobulin homology <IM2>
 F:218-238/Domain: transmembrane #status predicted <TM>
 F:239-650/Domain: intracellular #status predicted <INT>
 F:313-598/Domain: protein kinase homology <KIN>
 F:321-329/Region: protein kinase ATP-binding motif
 F:104,136,157,188/Binding site: carboxyrate (Asn) (covalent) #status predicted
 F:351,368,460/Active site: Lys, Glu, Asp #status predicted
 F:465,478/Binding site: magnesium (Asn, Asp) #status predicted
 F:491/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted

Query Match 41.9%; Score 44; DB 1; Length 650;
 Best Local Similarity 37.5%; Pred. No. 34;

Matches 9; Conservative 3; Mismatches 4; Indels 8; Gaps 1;

OY 1 PVSVCFF-----VGPDSFPD 16

Db 320 PLEGCGCGVCAALGMDSPRD 343

RESULT 12

T3185
 hypothetical protein F22F7.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33185

R:Miller, N.; Kramer, J.; Smith, A.

Submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid F22F7.

A:Reference number: 221299

A:Accession: T33185

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-313 <ML>

A:Cross-references: EMBL:AF067937; PIDN:MAC19217.1; GSPDB:GN00023; CESP:F22F7.6

A:Experimental source: strain Bristol NZ; clone F22F7

C:Genetics:

A:Gene: CESP:F22F7.6

A:Map position: 5

A:Introns: 18/2; 82/3; 117/1; 210/2; 241/3; 287/2

Query Match

Best Local Similarity 41.0%; Score 43; DB 2; Length 313;
 Pred. No. 23;

Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 PVSVCFFGPDASE 14

Db 51 PISVGIFIDFHSSQ 64

RESULT 13

F6185
 hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002

C:Accession: F6185

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Matzall,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; MUID:21016719; PMID:11130712

A:Accession: F6185

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <STO>
 A:Cross-references: GB:AE005172; NID:g238583; PIDN:AAB71464.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Escherichia coli probable zinc proteinase yael

Query Match

Best Local Similarity 41.0%; Score 43; DB 2; Length 441;
 Pred. No. 33;

Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 VCGFVGPDASEPPS 17

Db 139 LGGFVGFPPNDPDS 152

RESULT 14

T02547
 hypothetical protein Atg32480 [imported] - Arabidopsis thaliana

N:Alternate names: hypothetical protein T26B15.4

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001

C:Accession: T02547; F84733

R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke

submitted to the EMBL Data Library, July 1998

A:Description: Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.

A:Reference number: 214678

A:Accession: T02547

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-447 <KOU>

A:Cross-references: EMBL:AC004681; NID:g3298532; PID:g3298536

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: AB4420; MUID:20083487; PMID:10617197

A:Accession: F84733

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-447 <STO>

A:Cross-references: GB:AE002093; NID:g3298536; PIDN:AC25930.1; GSPDB:GN00139

C:Genetics:

A:Gene: T26B15.4; Atg32480

A:Map position: 2

C:Superfamily: Escherichia coli probable zinc proteinase yael

Query Match

Best Local Similarity 41.0%; Score 43; DB 2; Length 447;
 Pred. No. 34;

Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 VCGFVGPDASEPPS 17

Db 145 LGGFVGFPPNDPDS 158

RESULT 15

S59898
 kynureninase (EC 3.7.1.3) - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997

C:Accession: S59898

R:Takauchi, F.; Tsubouchi, R.; Yoshino, M.; Shibata, Y.

Biochim. Biophys. Acta 1252, 185-188, 1995

A>Title: Amino acid sequence of rat liver kynureninase.

A:Reference number: S59898; MUID:96049498; PMID:7578221

A:Accession: S59898

A>Status: preliminary

A:Molecule type: protein

A:Residues: 1-464 <TAK>

A:Keywords: hydrolase

Query Match 41.0%; Score 43; DB 2; Length 464;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 GCEVGF 11
 Db 244 GCEVGF 250

Search completed: March 27, 2003, 10:06:50
 Job time : 8.44512 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 : Search time 3.12805 Seconds
(without alignments)
251,930 Million cell updates/sec

Title: US-09-675-208-2
Perfect score: 105
Sequence: 1 PVSVCFCVGFDPASEPDSRH 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	41.9	112	1	HYPL_AGABI
2	44	41.9	428	1	TERM_DROME
3	43	41.0	464	1	KYNU_RAT
4	43	41.0	478	1	KYNU_CAEBL
5	42	40.0	364	1	Y3H1_ANASP
6	41.5	39.5	222	1	GL19_ARATH
7	41	39.0	5217	1	HHS1_COCCA
8	40	38.1	104	1	VE7_HPV32
9	40	38.1	282	1	POR1_BOVIN
10	40	38.1	282	1	POR1_RABIT
11	40	38.1	296	1	POR1_MOUSE
12	40	38.1	367	1	PGK_PAPPR
13	40	38.1	465	1	KYNU_HUMAN
14	40	38.1	615	1	ALBU_CHICK
15	40	38.1	650	1	PEX8_PICAN
16	40	38.1	755	1	AMO_KLEAE
17	40	38.1	834	1	DPOL_THECA
18	40	38.1	834	1	DPOL_THERM
19	40	38.1	893	1	DSCI_BOVIN
20	39.5	37.6	329	1	E2B1_MERYA
21	39.5	37.6	405	1	VGEG_HSV4
22	39.5	37.6	415	1	RL3_DROME
23	39	37.1	250	1	DAPB_CLOAB
24	39	37.1	269	1	THIG_CAUCR
25	39	37.1	301	1	MCAT_HUMAN
26	39	37.1	331	1	DEFA_MOUSE
27	39	37.1	348	1	YK23_CAEBL
28	39	37.1	389	1	BMRI_BACSU
29	39	37.1	481	1	KRI_HSV11
30	39	37.1	725	1	VG14_YEAST
31	39	37.1	752	1	8511_TRYCR
32	39	37.1	757	1	AMO_ECOLI
33	39	37.1	1088	1	RRPO_ROTBR

34	39	37.1	1088	1	RRPO_ROTBR	P21615 bovine rola
35	39	37.1	1088	1	RRPO_ROTBR	P17699 porcine rot
36	39	37.1	1088	1	RRPO_ROTBR	P22678 simian 11 r
37	39	37.1	1337	1	P152_YEAST	P39685 saeccharomyc
38	38.5	36.7	218	1	VG14_HSV4	P28941 equine herp
39	38.5	36.7	506	1	CE4B_RABIT	P15128 oryctolagus
40	38.5	36.7	1074	1	SM5A_MOUSE	O13591 homo sapien
41	38.5	36.7	1077	1	SM5A_MOUSE	O62217 mus musculu
42	38	36.2	111	1	VE7_HPV07	P36816 human papill
43	38	36.2	216	1	CSGD_SALTY	O54294 salmonella
44	38	36.2	301	1	MCAT_MOUSE	O94226 mus musculu
45	38	36.2	301	1	MCAT_RAT	P97521 rattus norv

ALIGNMENTS

RESULT 1
HYPL_AGABI STANDARD: PRT: 112 AA.
AC P49072:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrophobin 1 precursor (Hydrophobin A).
GN HYPA OR ABH1.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Horst H39;
RX MEDLINE=96192085; PubMed=8632464;
RA de Groot P.W.J., Schaap P.J., Visser J., van Griensven L.J.L.D.:
RT "The Agaricus bisporus hypha gene encodes a hydrophobin and
RT specifically accumulates in peel tissue of mushroom caps during fruit
RT body development."
RL J. Mol. Biol. 257:1008-1018(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Horst U3;
RX MEDLINE=96254018; PubMed=8704971;
RA Lugones L.G., Boesscher J.S., Scholtmeyer K., de Vries O.M.H.,
RA Wessels J.G.H.:
RT "An abundant hydrophobin (ABH1) forms hydrophobic rodlet layers in
RT Agaricus bisporus fruiting bodies."
RL Microbiology 142:1321-1329(1996).
CC - FUNCTION: FORMS HYDROPHOBIC RODLET LAYERS.
CC - SUBCELLULAR LOCATION: Secreted.
CC - DEVELOPMENTAL STAGE: EXPRESSED IN FRUITING BODIES.
CC - SIMILARITY: BELONGS TO THE FUNGAL HYDROPHOBIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X89242; CAA61530.1; -
CC EMBL: X90818; CAA62331.1; -
CC EMBL: X92861; CAA63447.1; -
CC InterPro: IPR001338; Hydrophobin.
CC Pfam: PF01185; Hydrophobin.
CC SMART: SM00075; HYDRO. 1.
CC PROSITE: PS00956; HYDROPHOBIN; 1.
KW Fruiting body; Cell wall; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 112 HYDROPHOBIN 1.
FT CONFLICT 40 41 CD -> Y (IN REF. 2).
FT SEQUENCE 112 AA; 11190 MW; F7647C12826FP6E1 CRC64;

```

Query Match          41.9%; Score 44; DB 1; Length 112;
Best Local Similarity 46.2%; Pred. No. 2.3;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 PVSVCFCVCFDAS 13
   1::11111:
Db 60 PINTGAFIAGFDCOT 72

RESULT 2
TERM DROME          STANDARD; PRT; 428 AA.
AC P11455; Q9VW03;
DT 01-OCT-1989 (Rel. 12, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Terminus protein.
GN TERM OR TER OR CG4216.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=805584; PubMed=3334721;
RA Balarelli R.M., Mahoney P.A., Salas F., Gustavson E., Boyer P.D.,
RA Chang M.-F., Roark M., Lengyel J.A.;
RT "Transcripts of the Drosophila blastoderm-specific locus, terminus,
RT finger."
RT Dev. Biol. 125:85-95(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adamats D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brockett P., Broctier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Fabios B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster G., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Jatali B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazetko M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```

```

RT "The genome sequence of Drosophila melanogaster.";
RC Science 287:2185-2195(2000).
CC -I- DEVELOPMENTAL STAGE: EXPRESSED FROM THE CELLULAR BLASTODERM STAGE
CC ON, MOST DURING GASTRULATION AND IS NO LONGER DETECTED BY THE END
CC OF GERM BAND EXTENSION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, M19140; AAA28928.1; -.
CC EMBL; AE003520; AAF49257.1; -.
CC PIR: A43741; A43741
CC FLYBASE; FBgn0003683; term.
CC InterPro; IPR000822; Znf_C2H2.
CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; UNKNOWN_1.
CC DNABinding; 325 346
CC ZN_FING 325 346
CC FT CONFLICT 144 144 A -> V (IN REF. 1)
CC SQ SEQUENCE 428 AA; 49165 MW; F54563F738AE434 CRC64;

Query Match          41.9%; Score 44; DB 1; Length 428;
Best Local Similarity 50.0%; Pred. No. 9;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 CRYGFPASPDSSRH 19
   1: 1111111
Db 37 CYSYSDADEPPSOH 50

RESULT 3
KYNUT_RAT
ID KYNUT_RAT          STANDARD; PRT; 464 AA.
AC P70712; Q9GW90;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE kynureninase (EC 3.7.1.3) (L-kynurenine hydrolase).
GN KYNUT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_Taxid=10116;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC STRAIN=Mistar; TISSUE=Liver;
RX MEDLINE=96049498; PubMed=7578221;
RA Takeuchi F., Tsubouchi R., Yoshino M., Shibata Y.;
RT "Amno-acid sequence of rat liver kynureninase."
RT Biochim. Biophys. Acta 1252:185-188(1995).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=97324088; PubMed=9180257;
RA Toma S., Nakamura M., Tone S., Okuno E., Kido R., Breton J.,
RA Avanzi N., Cozzi L., Speciale C., Mostardini M., Gatti S., Benatti L.;
RT "Cloning and recombinant expression of rat and human kynureninase."
RT FEBS Lett. 408:5-10(1997).
RN [3]
RP SEQUENCE OF 19-117 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver, and kidney;
RX MEDLINE=96314506; PubMed=8706755;
RA Alberici-Giani D., Buchli R., Malherbe P., Broger C., Lang G.,
RA Koehler C., Lahm H.-W., Cesura A.M.;
RT "Isolation and expression of a cDNA clone encoding human
RT kynureninase."
RT Eur. J. Biochem. 239:460-468(1996).
CC -I- FUNCTION: CATALYZES THE CLEAVAGE OF L-KYNURENINE (L-KYN) AND L-3-
CC HYDROXYKYNURENINE (L-3OHKYN) INTO ANTHRANILIC (AA) AND 3-

```

CC HYDROXYANTHRANILIC ACIDS (3-OHAA), RESPECTIVELY. HAS A PREFERENCE
CC FOR THE L-3-HYDROXY FORM. OPTIMUM ACTIVITY IS AROUND PH 9.0 FOR L-
CC KYN AND AROUND 8.5 FOR L-3OHKYN. ALSO HAS CYSTEINE-CONJUGATE-
CC BETA-LYASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: L-tryptophan + H(2)O = anthranilate + L-
CC alanine.
CC -1- CATALYTIC ACTIVITY: L-3-hydroxytryptophan + H(2)O = 3-
CC hydroxyanthranilate + L-alanine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- ENZYME REGULATION: INHIBITED BY O-METHYLBENZOTYALANINE (OMBA).
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF NAD COFACTORS FROM
CC TRYPTOPHAN THROUGH THE TRYPTOPHAN PATHWAY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN LIVER AND KIDNEY. ALSO
CC DETECTED IN HEART, RETINA, OVARY, LUNG, TESTIS AND BRAIN.
CC -1- INDUCTION: INHIBITED BY THIOL REAGENTS AND HEAVY METAL IONS.
CC -1- SIMILARITY: BELONGS TO THE TRYPTOPHANASE FAMILY. SLIGHTLY RELATED
CC TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U06168; AAC53206.1; -.
KW Hydroxylase; Pyridoxal phosphate; Acetylation.
FT MOD_RES 1 ACETYLATION.
FT BINDING 1 PYRIDOXAL PHOSPHATE.
FT CONFLICT 18 A -> T (IN REF. 1).
FT CONFLICT 26 D -> N (IN REF. 1).
FT CONFLICT 118 T -> S (IN REF. 1).
SO SEQUENCE 464 AA; 52453 MW; FFI035E3202ECC CRC64;

Query Match 41.0%; Score 43; DB 1; Length 464;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GCFVGF 11
Db 244 GCFVGF 250

RESULT 4
KYN CAEEL STANDARD; PRT; 478 AA.
ID KYN CAEEL
AC Q18026;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable kynureninase (EC 3.7.1.3) (L-kynurenine hydrolase).
GN C15H9.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Bentley D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Kohara Y., Shin'ichi T., Suzuki Y., Sugano S., Potdevin M.,
RA Thierry-Mieg Y., Thierry-Mieg D., Thierry-Mieg J.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-kynurenine + H(2)O = anthranilate + L-
CC alanine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE TRYPTOPHANASE FAMILY. SLIGHTLY RELATED
CC TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U56965; AAB52668.1; -.
DR EMBL: AF303267; AAC50225.1; -.
DR Wormpep: C15H9.7; CE06835.
KW Hypothetical protein; Hydroxylase; Pyridoxal phosphate.
FT BINDING 289 289 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SO SEQUENCE 478 AA; 54049 MW; E889450929EC94BD CRC64;

Query Match 41.0%; Score 43; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 GCFVGF 11
Db 257 GCFVGF 263

RESULT 5
Y3H1 ANASP STANDARD; PRT; 364 AA.
ID Y3H1 ANASP
AC Q8Y064;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc metalloprotease A113971 (EC 3.4.24.-).
GN A113971.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Matanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -1- COFACTOR: zinc (Probable).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50B.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP003594; BAB75670.1; -.
DR MEROPS: M50.004; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR004387; Zn_Mprotease.
DR InterPro: IPR000130; Zn_MTPpeptidase.
DR Pfam: PF00595; PDZ_1.
DR SMART: SM00228; PDZ_1.
DR TIGRFAWS: TIGR00054; mem_zinc_metalprot. 1.

DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KM Hypothetical protein: Hydrolase: Metalloprotease; Zinc: Transmembrane;
 KM Inner membrane: Complete proteome.
 FT METAL 17 17 ZINC (CATALYTIC) (POTENTIAL).
 FT ACT_SITE 18 18 POTENTIAL.
 FT METAL 21 21 ZINC (CATALYTIC) (POTENTIAL).
 FT TRANSMEM 92 114 POTENTIAL.
 FT TRANSMEM 281 303 POTENTIAL.
 FT TRANSMEM 329 346 POTENTIAL.
 FT DOMAIN 103 188 PDZ.
 SQ SEQUENCE 364 AA; 38613 MW; 54F6AE818AEFBEA CRC64;
 Query Match 40.0%; Score 42; DB 1; Length 364;
 Best Local Similarity 57.1%; Pred. No. 16;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 4 VCGFVGFDASEPDS 17
 DB 60 LGFVGFPDDPDS 73
 RESULT 6
 GLI9_ARATH STANDARD; PRT; 222 AA.
 ID GLI9_ARATH
 AC Q9FMB0;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Putative germin-like protein subfamily 1 member 9 precursor.
 GN At5G38910 OR K156.14 OR K156.90.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliopsida; euclidyales; core euclids; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 OX
 RN
 RX SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsuno A., Murai A., Nakayama S.,
 RA Nakazaki N., Nanno K., Okumura S., Shino S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spielgel L., Gao J., O'Shaughnessy A., Preston R.,
 RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
 RA Stoneking T., Pepin K., Spielth J., Sekhon M., Armstrong J., Becker M.,
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozerisky P., Riley A., Stromwall C.,
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Maria M.,
 RA Marienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Volckaert G., Wandurt R., Duesterhoft A., Stiekema W., Pohl T.,
 RA Enlhan K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
 RA Rampeger U., Medler H., Balke K., Medler E., Peters S.,
 RA van Staveren M., Dirkse W., Moolman P., Klein Iankhofst R.,
 RA Welzenegger T., Bothe G., Rose M., Hauf J., Berneriset S., Hempel S.,
 RA Feldpausch M., Lamberth S., Villarroel R., Gietlen J., Ardiles W.,
 RA Bents O., Lemcke K., Kolesov G., Mayer K., Rued S., Schoof H.,
 RA Scheller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 thaliana".
 RL Nature 408:823-826(2000).
 RN
 RN
 RC SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:41-54(1998).
 CC
 CC -1- FUNCTION: MAY PLAY A ROLE IN PLANT DEFENSE. HAS PROBABLY NO
 CC OXALATE OXIDASE ACTIVITY EVEN IF THE ACTIVE SITE IS CONSERVED.
 CC -1- SUBUNIT: OLIGOMER (BELIEVED TO BE A PENTAMER BUT PROBABLY
 CC HEXAMER) (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Apoplast (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB009048; BAB08648.1; -
 CC HSSP; P45850; 1F12.
 DR InterPro: IPR001929; Germin.
 DR Pfam: PF01072; Germin. 1.
 DR PRINTS; PR00325; GERMIN.
 DR PROSITE; PS00725; GERMIN. 1.
 KW Apoplast; Cell wall; Signal; Glycoprotein; Manganese; Metal-binding;
 KW Multigene family; Hypothetical protein.
 FT SIGNAL 1 22
 FT CHAIN 23 222
 FT METAL 111 111 SUBFAMILY 1 MEMBER 9.
 FT METAL 113 113 MANGANESE (BY SIMILARITY).
 FT METAL 118 118 MANGANESE (BY SIMILARITY).
 FT METAL 159 159 MANGANESE (BY SIMILARITY).
 FT DISULFID 32 49 BY SIMILARITY.
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 222 AA; 23884 MW; BB925CC41A905A4E CRC64;
 Query Match 39.5%; Score 41.5; DB 1; Length 222;
 Best Local Similarity 55.6%; Pred. No. 12;
 Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;
 QY 2 VSYG-CFVGFDASEPDS 18
 DB 122 VAVGTLEVFVSNPENR 139
 RESULT 7
 HTS1_COCOA STANDARD; PRT; 5217 AA.
 ID HTS1_COCOA
 AC Q01866;
 DT 01-FEB-1994 (Rel. 28; Created)
 DT 01-FEB-1994 (Rel. 28; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE HC-toxin synthetase (EC 6.3.2.-) (HTS).
 GN HTS1.
 OS Cochliobolus carbonum (Bipolaris zeicola).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; Cochliobolus.
 OC NCBI_TaxID=5017;
 OX
 RN
 RN
 RC SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 90305 / SB111;
 RC MEDLINE=93100328; PubMed=1281482;
 RA Scott-Craig J.S., Panaccione D.G., Pocard J.-A., Walton J.D.;
 RT "The cyclic peptide synthetase catalyzing HC-toxin production in the
 RT filamentous fungus Cochliobolus carbonum is encoded by a
 RT 15.7-kilobase open reading frame".
 RL J. Biol. Chem. 267:26044-26049(1992).
 RN
 RN
 RC FUNCTION.
 RC STRAIN=ATCC 90305 / SB111;
 RX MEDLINE=20138231; PubMed=10671527;

RA Cheng Y.-Q., Walton J.D.;
 RT "A eukaryotic alanine racemase gene involved in cyclic peptide
 biosynthesis";
 RL J. Biol. Chem. 275:4906-4911(2000).
 CC -1- FUNCTION: Non-ribosomal peptide synthetase, able to activate
 CC proline and Aeo (2-amino-9,10-epoxy-8-oxodecanoic acid), and
 CC epimerize L-Pro. Catalyzes the production of HC-toxin: a cyclic
 CC tetrapeptide. Activates and thioesterifies L-Pro, and epimerizes
 CC it to D-Pro; also uses D-Ala as a substrate but this is epimerized
 CC from L-Ala by TOXG.
 CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: CONTRAINS 4 ACYL CARRIER DOMAINS.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL: M98024; AAA33023.1; -.
 CC DR HSSP: P14687; 1AMU.
 CC DR InterPro: IPR000873; AMP-blind.
 CC DR InterPro: IPR001242; Condensatn.
 CC DR InterPro: IPR003880; Pantane_attach.
 CC DR Pfam: PF00501; AMP-binding; 4.
 CC DR Pfam: PF00550; pp-binding; 4.
 CC DR Pfam: PF00668; Condensation; 5.
 CC DR PRINTS: PR00154; AMPBINDING.
 CC DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 4.
 CC DR PROSITE: PS00455; AMP-BINDING; 3.
 CC DR PROSITE: PS50075; ACP_DOMAIN; 4.
 CC DR Ligase: Multifunctional enzyme: Phosphopantetheine; Repeat.
 CC FT REPEAT 249 842 DOMAIN 1.
 CC FT REPEAT 1854 2452 DOMAIN 2.
 CC FT REPEAT 3006 3606 DOMAIN 3.
 CC FT REPEAT 4158 4738 DOMAIN 4.
 CC FT DOMAIN 762 840 ACYL CARRIER (ACP) 1.
 CC FT DOMAIN 2384 2450 ACYL CARRIER (ACP) 2.
 CC FT DOMAIN 3536 3604 ACYL CARRIER (ACP) 3.
 CC FT DOMAIN 4667 4736 ACYL CARRIER (ACP) 4.
 CC FT BINDING 803 803 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC FT BINDING 2414 2414 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC FT BINDING 3568 3568 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC FT BINDING 4700 4700 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 CC FT BINDING 5217 5217 MW: 0331D9C5400163A5 CRC64;
 CC SO SEQUENCE

Query Match 39.0%; Score 41; DB 1; Length 5217;
 Best Local Similarity 57.1%; Pred. No. 3.6e+02;
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 5 GCFVGFDAEPPSR 18
 DB 4200 GCFVGFDAEPPSR 4213

RESULT 8
 VE7_HPV32
 ID VE7_HPV32 STANDARD: PRT: 104 AA.
 AC P36827;
 DT 01-JUN-1994 (rel. 29, Created)
 DT 01-JUN-1994 (rel. 29, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE E7 protein.
 GN E7.
 OS Human papillomavirus type 32.
 OC Viruses; dsDNA viruses, no RNA stage: Papillomaviridae;
 OC Papillomavirus.

OX NCBI_TaxID=10612;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94265501; PubMed=8205838;
 RA Delius H., Holmann B.;
 RT "Primer-directed sequencing of human papillomavirus types";
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC -1- FUNCTION: E7 PROTEIN HAS BOTH TRANSFORMING AND TRANS-ACTIVATING
 CC ACTIVITIES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL: X74475; CAA52550.1; -.
 CC DR PIR: S36510; S36510.
 CC DR InterPro: IPR000148; Papv_E7.
 CC DR Pfam: PF00527; E7_1.
 CC DR Early protein; Transcription regulation; Oncogene;
 CC DNA-binding; Trans-acting factor.
 CC KW C-X-X-C MOTIF-1.
 CC FT SITE 64 67 C-X-X-C MOTIF-2.
 CC FT SITE 97 100 C-X-X-C MOTIF-2.
 CC SO SEQUENCE 104 AA; 11591 MW; F9DFC8C0B6D804FF CRC64;
 CC

Query Match 38.1%; Score 40; DB 1; Length 104;
 Best Local Similarity 43.8%; Pred. No. 9.8;
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 PYSVGFGEFDSSEP 16
 DB 28 PVDLYCFEFDTSDED 43

RESULT 9
 P0R1_BOVIN
 ID P0R1_BOVIN STANDARD: PRT: 282 AA.
 AC P43879;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Voltage-dependent anion-selective channel protein 1 (VDAC-1)
 DE (Plasma membrane porin) (Brain-derived voltage-dependent anion channel 1)
 DE (BRI-VDAC).
 GN VDAC1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=94119914; PubMed=7507248;
 RA Dermietzel R., Hwang T.-K., Buettner R., Hofer A., Dotzler E.,
 RA Kremer M., Deutzmann R., Thümmes F.P., Fishman G.I., Spray D.C.,
 RA Sliemers D.;
 RT "Cloning and in situ localization of a brain-derived porin that
 RT constitutes a large-conductance anion channel in astrocytic plasma
 RT membranes";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:499-503(1994).
 CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE
 CC AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF
 CC SMALL HYDROPHILIC MOLECULES; IT ADOPTS AN OPEN CONFORMATION AT LOW
 CC OR ZERO MEMBRANE POTENTIAL, AND A CLOSED CONFORMATION AT POTENTIALS
 CC ABOVE 30-40 mV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY
 CC WHEREAS THE CLOSED STATE IS CATION-SELECTIVE.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE OF MITOCHONDRIA AND PLASMA
 CC MEMBRANE.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN BRAIN ASTROCYTES.

```

CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X75068; CAA52962.1; -
CC InterPro: IPR001925; Porin_Euk.
CC Pfam: PF01459; Euk_porin.1.
CC PRINTS: PR00185; EUKARYTPORIN.
CC PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
CC KMW plasma; Porin; Acetylation.
CC FT INIT MET 0 0 ACETYLATION (BY SIMILARITY).
CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
CC SEQ 282 AA; 30694 MW; 2E6B205FCD7849 CRC64;

Query Match
Best local Similarity 38.1%; Score 40; DB 1; Length 282;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 VSVGCFVGDASEPDSR 18
Db 122 INLGCDVDFDIAGPSIR 138

RESULT 10
PORL_RABIT STANDARD; PRT; 282 AA.
ID PORL_RABIT
AC O9TJ15;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Voltage-dependent anion-selective channel protein 1 (VDAC-1) (Outer
DE mitochondrial membrane protein porin 1).
GN VDAC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Corneal endothelium;
RA Rae J.L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE
CC AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF
CC SMALL HYDROPHILIC MOLECULES; IT ADOPTS AN OPEN CONFORMATION AT LOW
CC OR ZERO MEMBRANE POTENTIAL AND A CLOSED CONFORMATION AT POTENTIALS
CC ABOVE 30-40 mV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY
CC WHEREAS THE CLOSED STATE IS CATION-SELECTIVE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE OF MITOCHONDRIA AND PLASMA
CC MEMBRANE (BY SIMILARITY).
CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF209725; AAF22835.1; -
CC InterPro: IPR001925; Porin_Euk.
CC Pfam: PF01459; Euk_porin.1.
CC PRINTS: PR00185; EUKARYTPORIN.
CC PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
CC PROSITE: PS00558; EUKARYOTIC_PORIN; 1.

```

```

KW Outer membrane; Porin; Mitochondrion; Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT BINDING 72 72 DICYLOHEXYLCARODIIMIDE.
SQ SEQUENCE 282 AA; 30609 MW; 59F77E96F108A298 CRC64;

Query Match
Best local Similarity 38.1%; Score 40; DB 1; Length 282;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 VSVGCFVGDASEPDSR 18
Db 122 INLGCDVDFDIAGPSIR 138

RESULT 11
PORL_MOUSE STANDARD; PRT; 296 AA.
ID PORL_MOUSE
AC O60932;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Voltage-dependent anion-selective channel protein 1 (VDAC-1) (mVDAC1)
DE (mVDAC5) (outer mitochondrial membrane protein porin 1) (Plasmalemma)
DE Porin).
GN VDAC1 OR VDAC5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM MT-VDAC1).
RC TISSUE=Brain;
RX MEDLINE=96301405; PubMed=8660977;
RA Sampson M.J., Lovell R.S., Craigen W.J.;
RT "Isolation, characterization, and mapping of two mouse mitochondrial
RT voltage-dependent anion channel isoforms.";
RN [2]
RP CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX MEDLINE=20202612; PubMed=10716730;
RA Buettner R., Papoutsoglou G., Seemes E., Spray D.C., Dermietzel R.;
RT "Evidence for secretory pathway localization of a voltage-dependent
RT anion channel isoform.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3201-3206 (2000).
CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE
CC AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF
CC SMALL HYDROPHILIC MOLECULES; IT ADOPTS AN OPEN CONFORMATION AT LOW
CC OR ZERO MEMBRANE POTENTIAL AND A CLOSED CONFORMATION AT POTENTIALS
CC ABOVE 30-40 mV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY
CC WHEREAS THE CLOSED STATE IS CATION-SELECTIVE.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL VDAC1 (MT-VDAC1) IN OUTER
CC MEMBRANE OF MITOCHONDRIA AND PLASMALEMMAL VDAC1 (PL-VDAC1) IN
CC PLASMA MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PL-VDAC1 (SHOWN HERE) AND MT-
CC VDAC1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS OF EXPRESSION DETECTED IN HEART,
CC KIDNEY, BRAIN, AND SKELETAL MUSCLE. NOT EXPRESSED IN TESTIS.
CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U30840; AAB47777.1; -
CC SWISS-2DPAGE: O60932; MOUSE.
CC MGD: MGI:106919; Vdac1.
CC InterPro: IPR001925; Porin_Euk.

```

```

DR Pfam: PF01459; Euk_porin; 1.
DR PRINTS: PR00185; EUKARYOTIC_PORIN.
DR PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
KW Outer membrane; Porin; Mitochondrion; Alternative splicing;
KW Acetylation.
FT MOD_RES 14 14 ACETYLATION (BY SIMILARITY).
FT BINDING 86 86 DICYCLOHEXYLCARBODIIMIDE (BY SIMILARITY).
FT VARSPIC 1 13 MISSING (IN ISOFORM MT-VDAC1).
SO SEQUENCE 296 AA; 32351 MW; C0710C1717063B32 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 296;
Best Local Similarity 41.2%; Pred. No. 29;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Oy 2 VSGGFGVGFDASEPDSR 18
Db 136 INLCGVDFDIAGFSIR 152

RESULT 12
PGK_PAPPR STANDARD: PRT; 367 AA.
AC 000869;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoglycerate kinase (EC 2.7.2.3) (Fragment).
GN PKC.
OS Parametium primaurella.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Penicillia;
OC Parametium.
OX NCBI_TaxID=5886;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearlman R.E.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate -> ADP + 3-phospho-D-glycerol phosphate.
CC -1- PATHWAY: Second phase of glycolysis; second step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF001849; AAB58241.1; -.
DR HSSP: P00560; 10PG.
DR InterPro: IPR001576; PKC.
DR Pfam: PF00162; PGK; 1.
DR PROSITE: PS00111; POLYMERASE_KINASE; PARTIAL.
KW transferase; kinase; glycolysis.
FT NON_TER 1 1
FT NON_TER 1 1
SO SEQUENCE 367 AA; 39876 MW; DEF64500127DE460 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 367;
Best Local Similarity 42.1%; Pred. No. 36;
Matches 8; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Oy 1 PVSVCFGVGFDASEPDSR 19
Db 259 PTFVCGTGLDASSPVALH 277

RESULT 13
KYNH_HUMAN STANDARD: PRT; 465 AA.
AC Q16719;

```

```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE kynureninase (EC 3.7.1.3) (L-kynurenine hydrolase).
GN KYNH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP TISSUE=Hepatoma.
RC MEDLINE=96314506; PubMed=8706755;
RA Alberati-Giani D., Buchli R., Malherbe P., Broger C., Lang G.,
RA Koehler C., Lahm H.-W., Cesura A.M.;
RT "Isolation and expression of a cDNA clone encoding human kynureninase."
RL Eur. J. Biochem. 239:468(1996).
RN [2]
RP SEQUENCE FROM N.A.
RP TISSUE=liver;
RC MEDLINE=97324088; PubMed=9180257;
RA Toma S., Nakamura M., Tone S., Okuno E., Kido R., Breton J.,
RA Avanzi N., Cozzi L., Speciale C., Mostardini M., Gatti S., Benatti L.;
RT "Cloning and recombinant expression of rat and human kynureninase."
RL FEBS Lett. 408:5-10(1997).
CC -1- FUNCTION: CATALYZES THE CLEAVAGE OF L-KYNURENINE (L-KYN) AND L-3-HYDROXYKYNURENINE (L-3OHKYN) INTO ANTHRANILIC (AA) AND 3-HYDROXYANTHRANILIC ACIDS (3-OHAA), RESPECTIVELY. HAS A PREFERENCE FOR THE L-3-HYDROXY FORM. ALSO HAS CYSTEINE-CONJUGATE-BETA-LYASE ACTIVITY (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-kynurenine + H(2)O -> anthranilate + L-alanine.
CC -1- CATALYTIC ACTIVITY: L-3-hydroxykynurenine + H(2)O -> 3-hydroxyanthranilate + L-alanine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- ENZYME REGULATION: INHIBITED BY O-METHOXYBENZOTYALANINE (OMBA).
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF NAD COFACTORS FROM TRYPTOPHAN THROUGH THE KYNURENINE PATHWAY.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES TESTED (HEART, BRAIN PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS). HIGHEST LEVELS FOUND IN PLACENTA, LIVER AND LUNG. EXPRESSED IN ALL BRAIN REGIONS.
CC -1- INDUCTION: INCREASED LEVELS IN SEVERAL CEREBRAL AND SYSTEMIC INFLAMMATORY CONDITIONS.
CC -1- SIMILARITY: BELONGS TO THE KYNURENINASE FAMILY. SLIGHTLY RELATED TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U57721; AAC50650.1; -.
DR Genew; HGNC:6469; KYNH.
DR MIM: 605197; -.
DR MIM: 236800; -.
KW Hydrolase; Pyridoxal phosphate; Acetylation.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT BINDING 276 276 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SO SEQUENCE 465 AA; 52351 MW; BDD136BE18C79EBB CRC64;

Query Match 38.1%; Score 40; DB 1; Length 465;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 5 GCFVGF 11

```

DB 244 GCYVGF 250

RESULT 14

ALBU_CHICK STANDARD: PRT: 615 AA.

AC P19121.

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Serum albumin precursor.

GN ALB

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OC NCBI_Taxid=9031;

OX [1]

RN SEQUENCE FROM N.A.

RP TISSUE=Liver;

RC Cassady A.I., Salkind C.K., Baverstock P., Wallace J.C.;

RL Submitted (JUL-1991) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE OF 1-28 FROM N.A.

RX MEDLINE=83161037; PubMed=6187737;

RA Hache R.J.G., Wiskocil R., Vasa M., Roy R.N., Lau P.C.K., Deeley R.G.;

RT "The 5' noncoding and flanking regions of the avian very low density apolipoprotein II and serum albumin genes. Homologies with the egg white protein genes.";

RL J. Biol. Chem. 258:4556-4564(1983).

RN [3]

RP SEQUENCE OF 19-30.

RX MEDLINE=78019943; PubMed=911327;

RA Rosen A.M., Geller D.M.;

RT "Chicken microsomal albumin: amino terminal sequence of chicken proalbumin.";

RL Biochem. Biophys. Res. Commun. 78:1060-1066(1977).

CC -I- FUNCTION: SERUM ALBUMIN, THE MAIN PROTEIN OF PLASMA, HAS A GOOD BINDING CAPACITY FOR WATER, CA⁺⁺, NA⁺, K⁺, FATTY ACIDS, HORMONES, BILIRUBIN AND DRUGS. ITS MAIN FUNCTION IS THE REGULATION OF THE COLLOIDAL OSMOTIC PRESSURE OF BLOOD.

CC -I- SUBCELLULAR LOCATION: Secreted.

CC -I- TISSUE SPECIFICITY: PLASMA.

CC -I- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.

CC -I- SIMILARITY: CONTAINS 3 ALBUMIN DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X60688; CAA43098.1; -

DR EMBL: V00381; CAA23680.1; -

DR PIR: S15571; ABCS.

DR HSSP: P02768; 1E7B.

DR InterPro: IPR000264; Serum_albumin.

DR Pfam: PF00273; transport_prot. 3.

DR PRINTS: PRO0802; SERUMALBUMIN

DR ProDom: PD002486; Serum_albumin; 1.

DR SMART: SM00103; ALBUMIN; 3.

DR PROSITE: PS00212; ALBUMIN; 3.

KW Plasma; Metal-binding; Lipid-binding; Albumin; Repeat; Signal;

KW Copper.

FT SIGNAL 1 18

FT PROPEP 19 23

FT CHAIN 24 615

FT DOMAIN 24 209

FT DOMAIN 216 401

FT DOMAIN 408 599

FT METAL 30 30

CC COPPER (BY SIMILARITY).

FT DISULFID 80 89 BY SIMILARITY.

FT DISULFID 102 118 BY SIMILARITY.

FT DISULFID 117 128 BY SIMILARITY.

FT DISULFID 152 197 BY SIMILARITY.

FT DISULFID 196 205 BY SIMILARITY.

FT DISULFID 228 274 BY SIMILARITY.

FT DISULFID 273 281 BY SIMILARITY.

FT DISULFID 293 307 BY SIMILARITY.

FT DISULFID 306 317 BY SIMILARITY.

FT DISULFID 344 389 BY SIMILARITY.

FT DISULFID 388 397 BY SIMILARITY.

FT DISULFID 420 466 BY SIMILARITY.

FT DISULFID 465 476 BY SIMILARITY.

FT DISULFID 489 505 BY SIMILARITY.

FT DISULFID 504 515 BY SIMILARITY.

FT DISULFID 542 587 BY SIMILARITY.

FT DISULFID 586 595 BY SIMILARITY.

FT CARBOHYD 500 500 N-LINKED (GLCNAC, .) (POTENTIAL).

FT CONFLICT 24 24 F -> M (IN REF. 3).

FT SEQUENCE 615 AA; 69918 MW; E59E4BCCAC066C6 CRC64;

SQ

Query Match 38.1%; Score 40; DB 1; Length 615;

Best Local Similarity 54.5%; Pred No. 60;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 CFVGFDPSEPD 16

DB 128 CFLSFVQSQPD 138

RESULT 15

PEX8_PICAN STANDARD: PRT: 650 AA.

AC 000925;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Peroxisomal matrix protein PER1 precursor (peroxin-8).

GN PEX8 OR PER1.

OS Pichia angusta (Yeast) (Hansenula polymorpha).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.

OX NCBI_Taxid=4905;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CBS 4732;

RX MEDLINE=95050945; PubMed=7962056;

RA Waterham H.R., Titorenko V.I., Halma P., Gregg J.M., Harder W., Veenhuis M.;

RT "The Hansenula polymorpha PER1 gene is essential for peroxisome biogenesis and encodes a peroxisomal matrix protein with both carboxy- and amino-terminal targeting signals.";

RL J. Cell Biol. 127:737-749(1994).

CC -I- FUNCTION: ESSENTIAL FOR PEROXISOME BIOGENESIS. MAY PLAY A ROLE IN TRIGGERING THE PROTEIN IMPORT COMPETENCE OF INDIVIDUAL PEROXISOMES. IT MAY INTERACT WITH PEX8 (PEX10).

CC -I- SUBCELLULAR LOCATION: Peroxisomal; matrix.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Z30206; CAA82928.1; -

DR PROSITE: PS00342; MICROBODIES_CTER; 1.

KW Peroxisome; Transit peptide.

FT TRANSIT 1 650

FT CHAIN 2 650

FT SITE 648 650 MICROBODIES MATRIX PROTEIN PER1.

FT SITE 74122 MW; A9AC354204F50CTD CRC64;

SQ SEQUENCE 650 AA; 74122 MW; A9AC354204F50CTD CRC64;

Query Match 38.1%; Score 40; DB 1; Length 650;
 Best Local Similarity 36.8%; Pred. No. 64;
 Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 PVSVCCEVGFDASEPPSRH 19
 ||:||||:|:
 DB 145 PVAVGCLLSVDSRNDYDRY 163

Search completed: March 27, 2003, 10:04:17
 Job time : 5.12805 secs


```

DE Raffinose synthase (EC 2.4.1.82).
GN RFS.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WUNDER VON KETVEDON; TISSUE=SEED;
RT Peterhauer T., Mach L., Mucha J., Richter A.;
RT "Molecular characterization of raffinose synthase from pea (Pisum
RT sativum L.) seeds.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ426475; CAD20127.2;
KW Transferrase; Glycosyltransferase.
SQ SEQUENCE 798 AA; 88717 MW; 8D3F3ED5BF617B7 CRC64;

Query Match 51.4%; Score 54; DB 10; Length 798;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCFVGFDPSEPSRH 19
Db 84 GCFVGFNTTEAKSHH 98

RESULT 3
09M4M7 PRELIMINARY; PRT; 779 AA.
AC 09M4M7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Putative seed imbibition protein.
GN SIP.
OS Persea americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. HASS; TISSUE=MESOCARP;
RC Zamorano J.P., Evans A.D., Dopico B., Lowe A.L., Wilson I.D.,
RA Merodio C., Grierson D.;
RT "Isolation and characterization of cDNAs for mRNAs regulated during
RT cold storage of avocado (Persea americana Mill.) fruit.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ133148; CAB7245.1;
SQ SEQUENCE 779 AA; 85368 MW; C3A8BA3160316785 CRC64;

Query Match 47.6%; Score 50; DB 10; Length 779;
Best Local Similarity 60.0%; Pred. No. 9.7;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCFVGFDPSEPSRH 19
Db 42 GCFGATASESESIH 56

RESULT 4
08U687 PRELIMINARY; PRT; 836 AA.
AC 08U687;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein Atu6048.
GN Atu6048 OR AGR_PTI_100.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Plasmid pRiC58.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.

```

```

OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boeve D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nestler E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Gurolo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houniel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markez B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Seear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009423; AAL46284.1;
DR EMBL; AE007931; AAK91006.1;
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 836 AA; 91964 MW; 3C61FD8318D3DF1 CRC64;

Query Match 46.7%; Score 49; DB 16; Length 836;
Best Local Similarity 41.2%; Pred. No. 15;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 3 SVGCFVGFDPSEPSRH 19
Db 687 NLGATYLGFDAREPVAHV 703

RESULT 5
09F285 PRELIMINARY; PRT; 295 AA.
AC 09F285;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE 3-phosphoserine phosphatase (Putative 3-phosphoserine
DE phosphatase).
GN F26116.2 OR AT1G18640.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RP Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chiu J., Choi E., Gonzalez A.,
RA Howng B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharsky N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carinici P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,

```



```

RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southey A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene AtUG18640 (GI:15221897)";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC026238; AAF98410.1; -.
DR EMBL; AY065351; AAL38792.1; -.
DR InterPro: IPR001454; Hlganase/hydrolase.
DR InterPro: IPR004469; SerB.
DR Pfam: PF00702; Hydrolase; 1.
DR TIGRFAMS; TIGR00338; serB; 1.
SQ SEQUENCE 295 AA: 32318 MW: F14C95E636F7745E CRC64;

Query Match 44.8%; Score 47; DB 10; Length 295;
Best Local Similarity 64.3%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 GCFVGFDASEPPSR 18
DB 210 GFLGFDENEPTSR 223

RESULT 6
O82796 PRELIMINARY; PRT; 295 AA.
AC O82796;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 3-PHOSPHOSERIN phosphatase.
GN PSP.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Ho C., Noji M., Satou K.;
RT "Molecular cloning and characterization of PSP.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB018409; BAA33807.1; -.
DR EMBL; AB018409; BAA33806.1; -.
DR InterPro: IPR001454; Hlganase/hydrolase.
DR InterPro: IPR004469; SerB.
DR Pfam: PF00702; Hydrolase; 1.
DR TIGRFAMS; TIGR00338; serB; 1.
SQ SEQUENCE 295 AA: 32302 MW: F14C95E636E6754E CRC64;

Query Match 44.8%; Score 47; DB 10; Length 295;
Best Local Similarity 64.3%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 GCFVGFDASEPPSR 18
DB 210 GFLGFDENEPTSR 223

RESULT 7
O8UF84 PRELIMINARY; PRT; 336 AA.
AC O8UF84;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein AtU1515.
GN ATU1515 OR AGR_C_2794.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE-21608550; PubMed-11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Boyce D. St.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krupan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-21608551; PubMed-11743194;
RA Goodner B., Hinkle G., Gatlung S., Miller N., Blanchard M.,
RA Guorillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houniel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Mollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Planagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cleio C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009110; AAL42520.1; -.
DR EMBL; AE008075; AAK87306.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 336 AA: 36937 MW: C06405E207BF643A CRC64;

Query Match 44.8%; Score 47; DB 16; Length 336;
Best Local Similarity 47.1%; Pred. No. 13;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 PVSVCFCVGFDASEPPS 17
DB 135 PVTVCRCIGVDQDEPET 151

RESULT 8
O9FND9 PRELIMINARY; PRT; 783 AA.
AC O9FND9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Raffinose synthase protein.
GN MPO12.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-COLUMBIA;
RX MEDLINE-98069011; PubMed-9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;

```

```

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bosser L., Jones T., Banh J., Carrinchi P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shin P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB006702: BAB11595.1; -
DR EMBL: AY062781: AAL32855.1; -
DR EMBL: AY081645: AAM10207.1; -
SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;

Query Match
Best Local Similarity 44.3%; Score 46.5; DB 10; Length 783;
Matches 10; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY 2 VSAGCFVGFDA-SEPD5RH 19
DB 62 VSAGSIFGFLDGEPRSH 80

RESULT 9
O8YFH2 PRELIMINARY; PRT; 483 AA.
AC O8YFH2:
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE GTP-binding protein.
GN BME11550.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxId=29459;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Delyecchio V.G., Kapratel V., Redkar R.J., Patra G., Muter C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyriades N., Overbeek R.
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009590; AAL52731.1; -
DR InterPro: IPR005289; GTP-binding_dom.
DR InterPro: IPR005217; MMR_HSR1.
DR InterPro: IPR005225; SMALL_GTP.
DR Pfam: PF01926; MMR_HSR1.1.
DR TIGRfams: TIGR00650; MG442; 2.
DR TIGRfams: TIGR00231; small_GTP. 2.
DR Complete proteome.
KW COMPLETE PROTEOME.
SQ SEQUENCE 483 AA; 53311 MW; 162411DB29AA266 CRC64;

Query Match
Best Local Similarity 43.8%; Score 46; DB 16; Length 483;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PVSVCVGFDA-SEPD5 17
DB 184 PAAVGAIVGDIEDPDA 200

RESULT 10
O8VEA8 PRELIMINARY; PRT; 199 AA.
AC O8VEA8:
DT 01-MAR-2002 (TRENBLrel. 20, Created)

```

```

DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical 22.5 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC019395; AAH19395.1; -
DR InterPro: IPR003579; GTPase_Rad.
DR InterPro: IPR001806; Ras_transfmg.
DR InterPro: IPR000834; Zn_cardipect.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PR00449; RASTRNSFRMG.
DR SMART: SM00175; RAB; 1.
DR PROSITE: PS00133; CARBOXYPEPT_2N_2; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 199 AA; 22502 MW; 33353842079CF6FC CRC64;

Query Match
Best Local Similarity 42.9%; Score 45; DB 11; Length 199;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SVSCVGFDA-SEPD5 17
DB 81 SDCCIAAFDVTDPES 95

RESULT 11
O9YHD2 PRELIMINARY; PRT; 757 AA.
AC O9YHD2:
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Nuclear calmodulin-binding protein (Fragment).
GN URP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxId=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Lodge A.P., Walsh A., McNamee C.J., Moss D.J.
RT "Identification of CHURP, a Nuclear Calmodulin-Binding Protein Related
RT to hnRNP-U."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF098788; AAC69888.1; -
DR InterPro: IPR003878; SPRY_domain.
DR InterPro: IPR003877; SPRY_receptoc.
DR Pfam: PF00622; SPRY; 1.
DR SMART: SM00449; SPRY; 1.
FT NON_TER
SQ SEQUENCE 757 AA; 84179 MW; 7FA04ED9E176D1D CRC64;

Query Match
Best Local Similarity 42.9%; Score 45; DB 13; Length 757;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 VGCVGFDA-SEPD 16
DB 342 IGCFADEASEER 354

RESULT 12
O9QVW8 PRELIMINARY; PRT; 202 AA.
AC O9QVW8:
DT 01-MAR-2002 (TRENBLrel. 20, Created)

```


Best Local Similarity 50.0%; Pred. No. 51;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 CFVGFDPSEDSRH 19
|: |||||:
Db 36 CYSIDADEPSSOH 49

RESULT 15

09LZ54
ID 09LZ54 PRELIMINARY; PRT; 498 AA.
AC 09LZ54;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical 55.5 kDa protein.
GN T2P11_70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162971; CAB85985.1; -.
DR InterPro; IPR002068; HSP20.
DR InterPro; IPR000834; ZN_cardoPept.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR PROSITE; PS01031; HSP20; 1.
KW Hypothetical protein.
SQ SEQUENCE 498 AA; 55530 MW; 0791437EB8817BE2 CRC64;

Query Match 41.9%; Score 44; DB 10; Length 498;
Best Local Similarity 70.0%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCFVGFDPSE 14
||| ||| |:
Db 132 GCFTGFDPKSD 141

Search completed: March 27, 2003, 10:05:59
Job time : 13.6585 secs


```

; FILE REFERENCE: Bb1192 US CIP
; CURRENT APPLICATION NUMBER: US/09/873,880
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/363,321
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: 60/094,839
; PRIOR FILING DATE: July 31, 1998
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-873-880-36

Query Match          41.0%; Score 43; DB 10; Length 296;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCFVGFDPDSR 18
1 : ||| ||| |||
Db 211 GEVAGFDPDPDSR 224

RESULT 3
US-09-738-626-4094
; Sequence 4094, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4094
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4094

Query Match          40.0%; Score 42; DB 9; Length 382;
Best Local Similarity 64.3%; Pred. No. 46;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSVGCFVGFDPDSR 15
1 : ||| ||| |||
Db 249 VTVGLFVVLDASSP 262

RESULT 4
US-09-924-256A-22
; Sequence 22, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
```

```

; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Xanthoparmelia cumberlandia
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa=unknown amino acid
US-09-924-256A-22

Query Match          39.0%; Score 41; DB 10; Length 212;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCVGFDPDSR 16
1 : ||| : ||| : |||
Db 22 PEDVGCYIGACATDYD 37

RESULT 5
US-09-924-256A-34
; Sequence 34, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Leptogium corniculatum
US-09-924-256A-34

Query Match          39.0%; Score 41; DB 10; Length 212;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCVGFDPDSR 16
1 : ||| : ||| : |||
Db 22 PEDVGCYIGACATDYD 37

RESULT 6
US-10-033-297-6
; Sequence 6, Application US/10033297
; Publication No. US20020187486A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Mast, Andrea L.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
```

Sequential Invasive Cleavages
NUMBER OF SEQUENCES: 163
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Medien & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/033,297
FILING DATE: 12-NOV-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/350,597
FILING DATE: 09-JUL-1999
APPLICATION NUMBER: US/08/823,516
FILING DATE: 24-MAR-1997
APPLICATION NUMBER: PCT/US97/01072
FILING DATE: 21-JAN-1997
APPLICATION NUMBER: US 08/759,038
FILING DATE: 02-DEC-1996
APPLICATION NUMBER: US 08/758,314
FILING DATE: 02-DEC-1996
APPLICATION NUMBER: US 08/756,386
FILING DATE: 29-NOV-1996
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-033-297-6
Query Match 38.1%; Score 40; DB 9; Length 834;
Best Local Similarity 61.5%; Pred. NO. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 7 FVGFDAEPDSRH 19
DB 64 FVFDKAKAPSRH 76
RESULT 7
US-10-081-806-6
Sequence 6, Application US/10081806
Publication No. US20020197623A1
GENERAL INFORMATION:
APPLICANT: Prudent, James R.
HALL, JEFF G.
LYAMICHEV, VICTOR I.
TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Medien & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/081,806
FILING DATE: 22-Feb-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/756,386
FILING DATE: <unknown>
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02564
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-081-806-6
Query Match 38.1%; Score 40; DB 9; Length 834;
Best Local Similarity 61.5%; Pred. NO. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 7 FVGFDAEPDSRH 19
DB 64 FVFDKAKAPSRH 76
RESULT 8
US-10-074-328-6
Sequence 6, Application US/10074328
Publication No. US20030013098A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
GROTELUESCHEN HALL, JEFF S.
LYAMICHEV, VICTOR
OLIVE, DAVID M.
PRUDENT, JAMES R.
TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCES BY
INVADER-DIRECTED CLEAVAGE
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MEDIEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/074,328
FILING DATE: 12-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/599,491
FILING DATE: 23-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: P-40,027
REFERENCE/DOCKET NUMBER: FORS-01802
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-074-328-6

Query Match 38.1%; Score 40; DB 9; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAKPSFRH 76

RESULT 9
US-09-940-244-6
Sequence 6, Application US/09940244
Publication No. US20030044796A1
GENERAL INFORMATION:
APPLICANT: Neri, Bruce P.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamichev, Victor
APPLICANT: Smith, Lloyd M.
TITLE OF INVENTION: Reactions on Dendrimers
FILE REFERENCE: FORS-06478
CURRENT APPLICATION NUMBER: US/09/940,244
CURRENT FILING DATE: 2002-05-06
NUMBER OF SEQ ID NOS: 422
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 834
TYPE: PRT
ORGANISM: Thermus thermophilus
US-09-940-244-6

Query Match 38.1%; Score 40; DB 9; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAKPSFRH 76

RESULT 10
US-09-940-925A-6
Sequence 6, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
APPLICANT: LYAMICHEV, VICTOR I.
OLIVE, DAVID M.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-940-925A-6

Query Match 38.1%; Score 40; DB 9; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAKPSFRH 76

RESULT 11
US-09-940-244-259
Sequence 259, Application US/09940244
Publication No. US20030044796A1
GENERAL INFORMATION:
APPLICANT: Neri, Bruce P.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamichev, Victor
APPLICANT: Smith, Lloyd M.
TITLE OF INVENTION: Reactions on Dendrimers
FILE REFERENCE: FORS-06478
CURRENT APPLICATION NUMBER: US/09/940,244
CURRENT FILING DATE: 2002-05-06
NUMBER OF SEQ ID NOS: 422
SOFTWARE: PatentIn version 3.1
SEQ ID NO 259
LENGTH: 836
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-940-244-259

Query Match 38.1%; Score 40; DB 9; Length 836;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
||| |
Db 66 FVFDKAKPSFRH 78


```
RESULT 12
US-09-777-430A-8
; Sequence 8, Application US/09777430A
; Patent No. US20020128465A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Allawi, Halim T.
; APPLICANT: Wayland, Sarah R.
; APPLICANT: Takova, Tsetska
; APPLICANT: Neir, Bruce P.
; TITLE OF INVENTION: Charge Tags and the Separation of Nucleic Acid Molecules
; FILE REFERENCE: FORS-04912
; CURRENT APPLICATION NUMBER: US/09/777,430A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 836
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-777-430A-8

Query Match          38.1%; Score 40; DB 10; Length 836;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 7 FVGFDASEPPDSRH 19
    |||||
Db 66 FVFDKAPSPFRH 78
```

```
RESULT 13
US-09-777-430A-11
; Sequence 11, Application US/09777430A
; Patent No. US20020128465A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Allawi, Halim T.
; APPLICANT: Wayland, Sarah R.
; APPLICANT: Takova, Tsetska
; APPLICANT: Neir, Bruce P.
; FILE REFERENCE: FORS-04912
; CURRENT APPLICATION NUMBER: US/09/777,430A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 836
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-777-430A-11
```

```
Query Match          38.1%; Score 40; DB 10; Length 836;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 7 FVGFDASEPPDSRH 19
    |||||
Db 66 FVFDKAPSPFRH 78
```

```
RESULT 14
US-09-777-430A-15
; Sequence 15, Application US/09777430A
; Patent No. US20020128465A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Allawi, Halim T.
```

```
; APPLICANT: Wayland, Sarah R.
; APPLICANT: Takova, Tsetska
; APPLICANT: Neir, Bruce P.
; TITLE OF INVENTION: Charge Tags and the Separation of Nucleic Acid Molecules
; FILE REFERENCE: FORS-04912
; CURRENT APPLICATION NUMBER: US/09/777,430A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-777-430A-15
```

```
Query Match          38.1%; Score 40; DB 10; Length 842;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 7 FVGFDASEPPDSRH 19
    |||||
Db 66 FVFDKAPSPFRH 78
```

```
RESULT 15
US-09-777-430A-20
; Sequence 20, Application US/09777430A
; Patent No. US20020128465A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Allawi, Halim T.
; APPLICANT: Wayland, Sarah R.
; APPLICANT: Takova, Tsetska
; APPLICANT: Neir, Bruce P.
; TITLE OF INVENTION: Charge Tags and the Separation of Nucleic Acid Molecules
; FILE REFERENCE: FORS-04912
; CURRENT APPLICATION NUMBER: US/09/777,430A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-777-430A-20
```

```
Query Match          38.1%; Score 40; DB 10; Length 842;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 7 FVGFDASEPPDSRH 19
    |||||
Db 66 FVFDKAPSPFRH 78
```

```
Search completed: March 27, 2003, 10:20:36
Job time : 5.40244 secs
```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:40 ; Search time 7.31707 Seconds
(without alignments)
120.634 Million cell updates/sec

Title: US-09-675-208-1
Perfect score: 181
Sequence: 1 FGMCTWDAFYLTVPQVIEGVRLVDGCG 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	100.0	30	4	US-08-846-234-1
2	181	100.0	784	4	US-08-846-234-5
3	60	33.1	132	1	US-08-100-744-4
4	60	33.1	132	1	US-08-284-784-4
5	60	33.1	132	2	US-08-854-811-4
6	60	33.1	140	5	PCT-US90-06962-1
7	60	33.1	145	1	US-07-830-330-1
8	60	33.1	145	1	US-08-187-780-3
9	60	33.1	145	2	US-08-478-485-3
10	60	33.1	145	2	US-08-478-486F-3
11	60	33.1	146	2	US-08-231-894A-11
12	60	33.1	146	2	US-08-231-894A-12
13	60	33.1	146	2	US-08-231-894A-13
14	60	33.1	146	2	US-08-231-894A-14
15	60	33.1	146	3	US-08-231-894A-15
16	60	33.1	146	3	US-09-105-678A-49
17	60	33.1	146	4	US-09-421-208-49
18	60	33.1	146	4	US-09-385-114-2
19	60	33.1	146	4	US-09-417-721-3
20	60	33.1	146	4	US-09-417-721-5
21	60	33.1	146	6	5464943-6
22	60	33.1	146	6	5464943-8
23	60	33.1	146	6	5464943-10
24	60	33.1	146	6	5464943-12
25	60	33.1	146	6	5464943-14
26	60	33.1	146	6	5464943-25
27	60	33.1	146	6	5464943-26

28	60	33.1	147	6	5175147-8	Patent No. 5175147
29	60	33.1	147	6	5314872-1	Patent No. 5314872
30	60	33.1	150	1	US-08-441-629-8	Sequence 8, Appl1
31	60	33.1	150	3	US-08-776-207-8	Sequence 8, Appl1
32	60	33.1	150	5	PCT-US95-09172-8	Sequence 8, Appl1
33	60	33.1	150	5	PCT-US95-09172-8	Sequence 8, Appl1
34	60	33.1	153	3	US-08-325-186-2	Sequence 24, Appl1
35	60	33.1	154	2	US-08-438-439C-24	Sequence 1, Appl1
36	60	33.1	154	3	US-08-325-186-1	Sequence 6, Appl1
37	60	33.1	154	5	PCT-US91-02186-6	Sequence 6, Appl1
38	60	33.1	155	1	US-07-959-369-6	Sequence 7, Appl1
39	60	33.1	155	1	US-07-959-369-7	Sequence 2, Appl1
40	60	33.1	155	1	US-08-023-757-2	Sequence 4, Appl1
41	60	33.1	155	1	US-08-023-757-4	Sequence 1, Appl1
42	60	33.1	155	1	US-07-842-177A-1	Sequence 2, Appl1
43	60	33.1	155	1	US-08-177-502-2	Sequence 4, Appl1
44	60	33.1	155	1	US-08-177-502-4	Sequence 10, Appl1
45	60	33.1	155	1	US-08-439-725A-10	

ALIGNMENTS

RESULT 1
US-08-846-234-1
Sequence 1, Application US/08846234
Patent No. 6166292

GENERAL INFORMATION:
APPLICANT: OSUMI Chleko
APPLICANT: NOZAKI Jinshi

APPLICANT: KIDA Takao
TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT

NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:

ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR

CITY: ARLINGTON
STATE: VIRGINIA

COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846,234
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618
TELEPHONE: (703)-413-3000

TELEFAX: (703)-413-2220
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids

TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
FRAGMENT TYPE: internal

US-08-846-234-1

Query Match 100.0%; Score 181; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 5.2e-20;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 FGMCTWDAFYLTVPQVIEGVRLVDGCG 30
1 FGMCTWDAFYLTVPQVIEGVRLVDGCG 30

RESULT 2
US-08-846-234-5
Sequence 5, Application US/08846234
Patent No. 6166292
GENERAL INFORMATION:
APPLICANT: OSUMI Chieko
APPLICANT: NOZAKI Jinshi
APPLICANT: KIDA Takao
TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,234
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-413-3000
TELEFAX: (703)-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 784 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-846-234-5
Query Match 100.0%; Score 181; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 2,7e-18;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 FGCWTDAFYLVHPQVIEGVRLVGGC 30
Db 215 FGCWTDAFYLVHPQVIEGVRLVGGC 244
RESULT 3
US-08-100-744-4
Sequence 4, Application US/08100744
Patent No. 5563046
GENERAL INFORMATION:
APPLICANT: MASCARENHAS, DESMOND
APPLICANT: ZHANG, SUNNY
APPLICANT: OLSEN, PAMELA
APPLICANT: OLSEN, DAVID
APPLICANT: CARRILLO, PEDRO A.
TITLE OF INVENTION: POLYPEPTIDE FUSIONS TO
TITLE OF INVENTION: INTERLEUKIN-1-LIKE POLYPEPTIDES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,784
FILING DATE: 02-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LUTHER, BARBARA J.
REGISTRATION NUMBER: 33,954
REFERENCE/DOCKET NUMBER: 22095-20275.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-100-744-4
Query Match 33.1%; Score 60; DB 1; Length 132;
Best Local Similarity 40.0%; Pred. No. 0.17;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
Qy 3 WCTWDAFYLVHPQVIEGVRLVLD 27
Db 10 YCKNGGFRLRHPDGRVDCGRKSD 34
RESULT 4
US-08-284-784-4
Sequence 4, Application US/08284784
Patent No. 5629172
GENERAL INFORMATION:
APPLICANT: MASCARENHAS, DESMOND
APPLICANT: ZHANG, YANG
APPLICANT: OLSEN, PAMELA S.
APPLICANT: OLSEN, DAVID R.
APPLICANT: CARRILLO, PEDRO A.
TITLE OF INVENTION: EXPRESSION OF FUSION POLYPEPTIDES
TITLE OF INVENTION: TRANSPORTED OUT OF THE CYTOPLASM WITHOUT LEADER SEQUENCES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,784
FILING DATE: 02-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: PARK, FREDIE K.
REGISTRATION NUMBER: 35,636
REFERENCE/DOCKET NUMBER: 22095-20275.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/100,744
FILING DATE: 02-AUG-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LUTHER, BARBARA J.
REGISTRATION NUMBER: 33,954
REFERENCE/DOCKET NUMBER: 22095-20275.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-100-744-4
Query Match 33.1%; Score 60; DB 1; Length 132;
Best Local Similarity 40.0%; Pred. No. 0.17;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
Qy 3 WCTWDAFYLVHPQVIEGVRLVLD 27
Db 10 YCKNGGFRLRHPDGRVDCGRKSD 34
RESULT 4
US-08-284-784-4
Sequence 4, Application US/08284784
Patent No. 5629172
GENERAL INFORMATION:
APPLICANT: MASCARENHAS, DESMOND
APPLICANT: ZHANG, YANG
APPLICANT: OLSEN, PAMELA S.
APPLICANT: OLSEN, DAVID R.
APPLICANT: CARRILLO, PEDRO A.
TITLE OF INVENTION: EXPRESSION OF FUSION POLYPEPTIDES
TITLE OF INVENTION: TRANSPORTED OUT OF THE CYTOPLASM WITHOUT LEADER SEQUENCES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,784
FILING DATE: 02-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: PARK, FREDIE K.
REGISTRATION NUMBER: 35,636
REFERENCE/DOCKET NUMBER: 22095-20275.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent InRelease #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/830,330
FILING DATE: 19920420
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Obilon, No. 5288704man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 769-230-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-07-830-330-1

Query Match	33.1%;	Score 60;	DB 1;	Length 145;
Best Local Similarity	40.0%;	Pred. No. 0.19;		
Matches 10;	Conservative 5;	Mismatches 10;	Indels 0;	Gaps 0;

```

QY      3 WCETDAFYLTVHPQGVIEGVRHLVD 27
      .  :| :| :| | :| | |
Db     24 YCKNGGFRLRHPDGRVDGVREKSD 48

```

RESULT 8
US-08-187-780-3
; Sequence 3, Application US/08187780
; Patent No. 5459250

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Diskette, 5.25 inch
3  MEDIUM TYPE: 360 Kb storage
4  COMPUTER: IBM or IBM-compatible
5  OPERATING SYSTEM: PC/MS-DOS
6  SOFTWARE: Wordperfect
7  CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/187,780
FILING DATE: January 25, 1994

CLASSIFICATION: 530
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/901,705
FILING DATE: June 22, 1992

APPLICATION NUMBER: 07/806,771
FILING DATE: December 6, 1991

APPLICATION NUMBER: 07/177,506
FILING DATE: April 4, 1988

APPLICATION NUMBER: 07/062,925
FILING DATE: June 16, 1987

ATTORNEY/AGENT INFORMATION:
NAME: Howard M. Frankfort

```

;      REGISTRATION NUMBER:  32,613
;
;      REFERENCE/DOCKET NUMBER:  5986/13586-US3

```

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
SEQUENCE: 145

```

;
; TYPE: amino acid
; STRANDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: Protein
; FEATURE:
;

```

NAME/KEY:
LOCATION:

IDENTIFICATION METHOD: TH
OTHER INFORMATION: TH

OTHER INFORMATION: corresponding to bovine basic fibroblast growth factor, can be found in Table 1,

OTHER INFORMATION: page 9, lines 9, 14, and 19, in the application, as filed.

; PUBLICATION INFORMATION:
;
; AUTHORS:
;
;

TITL:
JOUR:

VOLUME:
ISSUE:
;

PAGES: _____
DATE: _____
; _____
; _____

DOCUMENT NUMBER:
FILING DATE:
EXPIRATION DATE:

```

; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO: 1-145
;
PDB:08-187-790-3

```

Query Match	33.1%	Score 60;	DB 1;	Length 145;
Best Local Similarity	40.0%;	Pred. No. 0.19;		
Matches 10;	Conservative 5;	Mismatches 10;	Indels 0;	Gaps 0;

QY 3 WCFTDAFYLTVHPQGVIEGVRHLVD 27
:| | :| | :| | |
Db 24 YCKNGFFLRHPDGRVDGVREKSD 48

RESULT 9
US-08-478-485-3
; Sequence 3, Application US/08478485
; Patent No. 5883071

GENERAL INFORMATION:
APPLICANT: CLAUDIO BASILICO

APPLICANT: DANIELA TALARICO
TITLE OF INVENTION: MAMMALIAN GROWTH FACTOR

```

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Darby & Darby P.C.
STREET: 805 Third Avenue

CITY: New York
STATE: New York

COUNTRY: USA
ZIP: 10022

```

;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy Diskette, 3+ inch

```

```

; MEDIUM TYPE: 1.44 MB storage
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC/MS-D
; SOFTWARE: Wordperfect 5.1

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,485
;

```

```
;; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 424
```

;; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,780
;

FILING DATE: January 25, 1994
APPLICATION NUMBER: 07/901,705

FILING DATE: June 22, 1992

APPLICATION NUMBER: 07/806,771
FILING DATE: December 6, 1991
APPLICATION NUMBER: 07/177,506
FILING DATE: April 4, 1988
APPLICATION NUMBER: 07/062,925
FILING DATE: June 16, 1987
ATTORNEY/AGENT INFORMATION:
NAME: Joseph R. Robinson
REGISTRATION NUMBER: 33,448
REFERENCE/DOCKET NUMBER: 5986/13586-US6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 145
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
DESCRIPTION: Protein
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: This sequence,
OTHER INFORMATION: corresponding to bovine basic fibroblast
OTHER INFORMATION: growth factor, can be found in Table 1,
OTHER INFORMATION: page 9, lines 9, 14, and 19, in the
OTHER INFORMATION: application, as filed.
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 1-145
US-08-478-485-3
Query Match 33.1%; Score 60; DB 2; Length 145;
Best Local Similarity 40.0%; Pred. No. 0.19;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
OY 3 WCTMDAFYLVHPQGVIEGVRLVD 27
Db 24 YCKNGFRLRIHPDGRVDGVREKSD 48
RESULT 10
US-08-478-486F-3
Sequence 3, Application US/08478486F
Patent No. 6432702
GENERAL INFORMATION:
APPLICANT: CLAUDIO BASILICO
APPLICANT: DANIELA TALARICO
TITLE OF INVENTION: MAMMALIAN GROWTH FACTOR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby P.C.
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Diskette, 3+ inch,
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,486F
FILING DATE: June 7, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,780
FILING DATE: January 25, 1994
APPLICATION NUMBER: 07/901,705
FILING DATE: June 22, 1992
APPLICATION NUMBER: 07/806,771
FILING DATE: December 6, 1991
APPLICATION NUMBER: 07/177,506
FILING DATE: April 4, 1988
APPLICATION NUMBER: 07/062,925
FILING DATE: June 16, 1987
ATTORNEY/AGENT INFORMATION:
NAME: Howard M. Frankfort
REGISTRATION NUMBER: 32,613
REFERENCE/DOCKET NUMBER: 5986/13586-US7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 145
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
DESCRIPTION: Protein
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: This sequence,
OTHER INFORMATION: corresponding to bovine basic fibroblast
OTHER INFORMATION: growth factor, can be found in Table 1,
OTHER INFORMATION: page 9, lines 9, 14, and 19, in the
OTHER INFORMATION: application, as filed.
US-08-478-486F-3
Query Match 33.1%; Score 60; DB 4; Length 145;
Best Local Similarity 40.0%; Pred. No. 0.19;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
OY 3 WCTMDAFYLVHPQGVIEGVRLVD 27
Db 24 YCKNGFRLRIHPDGRVDGVREKSD 48
RESULT 11
US-08-231-894A-11
Sequence 11, Application US/08231894A
Patent No. 5851990
GENERAL INFORMATION:
APPLICANT: FUJISHIMA, AKIRA
APPLICANT: FUKUDA, TSUNEHIKO
TITLE OF INVENTION: BRGF MUTEDIN AND ITS PRODUCTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. COMLIN, DIKE, BRONSTEIN, ROBERTS
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

```
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,894A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/873907
FILING DATE: 24-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 097655-1991
FILING DATE: 26-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 066381-1992
FILING DATE: 24-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 41769-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-231-894A-11

Query Match          33.1%  Score 60:  DB 2:  Length 146:
Best Local Similarity 40.0%  Pred. No. 0.19;
Matches 10:  Conservative 5;  Mismatches 10;  Indels 0;  Gaps 0;

Qy      3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
      : | : | : | : | : | : | : | : |
Db      24 YCKNGGFRLRHDPGRVGDVREKSD 48

RESULT 12
US-08-231-894A-12
Sequence 12, Application US/08231894A
Patent No. 5851990
GENERAL INFORMATION:
APPLICANT: FUKUDA, AKIRA
APPLICANT: FUKUDA, TSUNEHIKO
TITLE OF INVENTION: BFGF MUTEIN AND ITS PRODUCTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
ADDRESS: 6 CUSHMAN
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,894A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/873907
FILING DATE: 24-APR-1992
```

```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 097655-1991
FILING DATE: 26-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 066381-1992
FILING DATE: 24-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 41769-FWC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 146 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-231-894A-12

Query Match          33.1%  Score 60:  DB 2:  Length 146:
Best Local Similarity 40.0%  Pred. No. 0.19;
Matches 10:  Conservative 5;  Mismatches 10;  Indels 0;  Gaps 0;

Qy      3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
      : | : | : | : | : | : | : | : |
Db      24 YCKNGGFRLRHDPGRVGDVREKSD 48

RESULT 13
US-08-231-894A-13
Sequence 13, Application US/08231894A
Patent No. 5851990
GENERAL INFORMATION:
APPLICANT: FUKUDA, AKIRA
APPLICANT: FUKUDA, TSUNEHIKO
TITLE OF INVENTION: BFGF MUTEIN AND ITS PRODUCTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
ADDRESS: 6 CUSHMAN
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,894A
FILING DATE: 22-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/873907
FILING DATE: 24-APR-1992
CURRENT APPLICATION DATA:
APPLICATION NUMBER: JP 097655-1991
FILING DATE: 26-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 066381-1992
FILING DATE: 24-MAR-1992
ATTORNEY/AGENT INFORMATION:
```



```

; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 41769-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 146 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; US-08-231-894A-13

```

```

Query Match 33.1%; Score 60; DB 2; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.19;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

```

```

OY 3 WCTWDAFLYVHPQGVICGVRHLVD 27
; 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 24 YCKNGFFLRHPDGRVDGVREKSD 48

```

```

RESULT 14
US-08-231-894A-14
; Sequence 14, Application US/08231894A
; Patent No. 5851990
; GENERAL INFORMATION:
; APPLICANT: FUJISHIMA, AKIRA
; TITLE OF INVENTION: BFGF MUTEIN AND ITS PRODUCTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,894A
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/873907
; FILING DATE: 24-APR-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 097655-1991
; FILING DATE: 26-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 066381-1992
; FILING DATE: 24-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 41769-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 146 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; US-08-231-894A-14

```

```

Query Match 33.1%; Score 60; DB 2; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.19;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

```

```

OY 3 WCTWDAFLYVHPQGVICGVRHLVD 27
; 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 24 YCKNGFFLRHPDGRVDGVREKSD 48

```

```

RESULT 15
US-08-231-894A-15
; Sequence 15, Application US/08231894A
; Patent No. 5851990
; GENERAL INFORMATION:
; APPLICANT: FUJISHIMA, AKIRA
; TITLE OF INVENTION: BFGF MUTEIN AND ITS PRODUCTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,894A
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/873907
; FILING DATE: 24-APR-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 097655-1991
; FILING DATE: 26-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 066381-1992
; FILING DATE: 24-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 41769-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 146 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal

```

Fri Apr 4 12:31:19 2003

us-09-675-208-1.rai

Page 8

ORIGINAL SOURCE:
US-08-231-894A-15

US-08-231-894A-15

Query Match	33.1%	Score 60;	DB 2;	Length 146;
Best Local Similarity	40.0%;	Pred NO	0.19;	

Matches	Conservative	5	Mismatches	10	Indels	0	Gaps	0
10	5	10	0	0	0	0	0	0

```

Oy      3 WCITDAFYLTVHPQGVIEGVRHLVD 27
          :| :| :| :| :| :| :| :| :|
Db      24 YCKNGGFELRIHPDGRVDGVREKSD 48

```

Db 24 YCKNGGFFLRHPDGRVDGVREKSD 48

Search completed: March 27, 2003, 10:07:33
Job time : 8.31707 secs

Job time : 8.31707 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 21.2195 Seconds

(without alignments)
188.389 Million cell updates/sec

Title: us-09-675-208-1

Perfect score: 181

Sequence: 1 FGWCTWDAFYLVHPGVIEGVRLVDGCG 30

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	100.0	30	19	AAW53567
2	181	100.0	30	20	AAV17422
3	181	100.0	784	19	AAW53570
4	181	100.0	784	20	AAV17417
5	158	87.3	758	21	AAV70978
6	148	81.8	799	19	AAW57886
7	148	81.8	799	19	AAV30142
8	146	80.7	587	19	AAW57888
9	146	80.7	587	20	AAV30144
10	146	80.7	780	22	AAW57859

11	146	80.7	781	19	AAW57887
12	146	80.7	781	20	AAV30143
13	146	80.7	781	22	AAW57888
14	144	79.6	572	20	AAV32075
15	144	79.6	577	20	AAV32074
16	144	79.6	783	23	ABW3664
17	122	67.4	783	20	AAV32073
18	108	59.7	751	21	AAV70980
19	107	59.1	763	21	AAV70977
20	106	58.6	750	20	AAV17418
21	106	58.6	807	23	ABW3665
22	102	56.4	770	21	AAV70976
23	98	54.1	841	21	AAV70981
24	97	53.6	756	21	AAV70975
25	71	39.2	10	20	AAV17419
26	68	37.6	150	11	AAW03961
27	62	34.3	40	14	AAW43287
28	60	33.1	40	14	AAW43286
29	60	33.1	43	13	AAW43442
30	60	33.1	44	11	AAW18542
31	60	33.1	45	8	AAW71559
32	60	33.1	45	14	AAW43278
33	60	33.1	45	21	AAW18551
34	60	33.1	86	9	AAW81933
35	60	33.1	101	10	AAW90557
36	60	33.1	105	10	AAW90558
37	60	33.1	114	10	AAW90559
38	60	33.1	114	22	AAW60696
39	60	33.1	118	10	AAW90560
40	60	33.1	121	23	AAW81292
41	60	33.1	123	9	AAW90561
42	60	33.1	129	10	AAW81940
43	60	33.1	129	10	AAW90562
44	60	33.1	129	10	AAW90564
45	60	33.1	132	20	AAV17995

ALIGNMENTS

RESULT 1
AAW53567
ID AAW53567 standard; peptide: 30 AA.
XX
AC AAW53567;
XX
DT 06-JUL-1998 (first entry)
XX
DE Cucumber raffinose synthase residues 215 to 244.
XX
KW Cucumber; raffinose synthase; sucrose; galactinol.
XX
OS Cucumis sativus.
XX
PN JPI0084973-A.
XX
PD 07-APR-1998.
XX
PF 28-APR-1997; 97JP-0111124.
XX
PR 26-JUL-1996; 96JP-0198079.
XX
PR 26-APR-1996; 96JP-0107682.
XX
PA (AJIN) AJINOMOTO KK.
XX
DR WPI: 1998-264858/24.
XX
PT Raffinose synthase gene - useful for preparation of raffinose in
XX transformed plant
XX
PS Claim 2: Page 16: 26pp; Japanese.
XX
CC The present sequence is a cucumber raffinose synthase fragment.

CC Raffinose synthase forms raffinose from sucrose and galactinol, has
 CC an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees
 CC C, has a molecular weight of 75 to 95 kDa by gel filtration or 90
 CC to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is
 CC inhibited by iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 30 AA;

Query Match 100.0%; Score 181; DB 19; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2.5e-19;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30
 |||
 DB 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30

RESULT 2

AAV17422
 ID AAV17422 standard; peptide; 30 AA.

XX AC AAV17422;

XX DT 29-JUL-1999 (first entry)

XX DE Cucumber raffinose synthase peptide SEQ ID NO:1.

XX KW Raffinose synthase; sucrose; galactinol.

XX OS Cucumis sativus.

XX PN JP1123080-A.

XX PD 11-MAY-1999.

XX PF 24-OCT-1997; 97JP-0292969.

XX PR 24-OCT-1997; 97JP-0292969.

XX PA (AJIN) AJINOMOTO KK.

XX DR WPI; 1999-340516/29.

XX PT New raffinose synthase gene - for production of raffinose from
 XX sucrose and galactinol

XX PS Example 2; Page 21; 37pp; Japanese.

XX CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents a raffinose
 CC synthase peptide from cucumber.

XX SQ Sequence 30 AA;

Query Match 100.0%; Score 181; DB 20; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2.5e-19;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30
 |||
 DB 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30

RESULT 3

AAW53570
 ID AAW53570 standard; Protein; 784 AA.

XX AC AAW53570;

XX DT 06-JUL-1998 (first entry)

XX DE Cucumber raffinose synthase.

XX KW Cucumber; raffinose synthase; sucrose; galactinol.

XX OS Cucumis sativus.

XX PN JP10084973-A.

XX PD 07-APR-1998.

XX PF 28-APR-1997; 97JP-0111124.

XX PR 26-JUL-1996; 96JP-0198079.

XX PR 26-APR-1996; 96JP-0107682.

XX PA (AJIN) AJINOMOTO KK.

XX DR WPI; 1998-264858/24.

XX DR N-PSDB; AAV22250.

XX PT Raffinose synthase gene - useful for preparation of raffinose in
 XX transformed plant

XX PS Claim 3; Pages 17-20; 26pp; Japanese.

XX CC The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC iodoacetamide, N-ethylmaleimide and myoinositol.

XX SQ Sequence 784 AA;

Query Match 100.0%; Score 181; DB 19; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1.2e-17;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30
 |||
 DB 215 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 244

RESULT 4

AAV17417
 ID AAV17417 standard; Protein; 784 AA.

XX AC AAV17417;

XX DT 29-JUL-1999 (first entry)

XX DE Cucumber raffinose synthase.

XX KW Raffinose synthase; sucrose; galactinol.

XX OS Cucumis sativus.

XX PN JP1123080-A.

XX PD 11-MAY-1999.

XX PF 24-OCT-1997; 97JP-0292969.

XX PR 24-OCT-1997; 97JP-0292969.

XX PA (AJIN) AJINOMOTO KK.

XX DR WPI; 1999-340516/29.

XX DR N-PSDB; AAX61238.

XX PT New raffinose synthase gene - for production of raffinose from
 XX sucrose and galactinol

XX Claim 2: Page 25-27; 37pp; Japanese.
 XX
 CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents raffinose
 CC synthase from cucumber.
 XX
 SQ Sequence 784 AA:
 Query Match 100.0%; Score 181; DB 20; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1,2e-17;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 FGMCTWDAFYLVHPGVIEGVRLVDGCG 30
 Db 215 FGMCTWDAFYLVHPGVIEGVRLVDGCG 244

RESULT 5
 AAY70978
 ID AAY70978 standard; Protein; 758 AA.
 AC AAY70978;
 XX
 DT 09-AUG-2000 (first entry)
 XX
 DE Soybean raffinose synthase from clone sfl1.pk125.d4.
 XX
 KM Soybean; raffinose synthase; raffinose saccharide;
 KM clone sfl1.pk125.d4; nutritional; soy protein.
 XX
 OS Glycine max.
 XX
 PM WO200024915-A2.
 PD 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99MO-US24923.
 XX
 PR 23-OCT-1998; 98US-0105451.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Allen SM, Hiltz WD;
 XX
 DR WPI: 2000-350754/30.
 DR N-PSDB: AAD00335.
 XX
 PT Nucleic acids and encoded proteins involved in the biosynthesis of
 PT raffinose, useful for producing soybean seeds with a reduced raffinose
 PT content and therefore improved nutritional quality -
 XX
 PS Claim 2: Page 47-49; 58pp; English.
 CC
 CC The present sequence is a raffinose synthase from
 CC clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
 CC library sfl1. Raffinose synthase is involved in the biosynthesis
 CC of raffinose and higher homologues in the raffinose saccharide family
 CC from sucrose. The present sequence is useful for reducing the raffinose
 CC saccharide content of soybean seeds which improves the nutritional
 CC quality of the soy protein products derived from them.
 XX
 SQ Sequence 758 AA:

Query Match 87.3%; Score 158; DB 21; Length 758;
 Best Local Similarity 86.7%; Pred. No. 2,6e-14;
 Matches 26; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 FGMCTWDAFYLVHPGVIEGVRLVDGCG 30
 Db 215 FGMCTWDAFYLVHPGVIEGVRLVDGCG 30

Db 202 FGMCTWDAFYLVHPGVIEGVRLVDGCG 231

RESULT 6
 AAY57886
 ID AAY57886 standard; Protein; 799 AA.
 AC AAY57886;
 XX
 DT 23-SEP-1998 (first entry)
 XX
 DE Broad bean raffinose synthetase.
 DE Raffinose synthetase; metabolism modification; food additive;
 DE gastrointestinal flora; broad bean.
 XX
 OS Vicia faba.
 XX
 PM EP849359-A2.
 PD 24-JUN-1998.
 XX
 PF 18-DEC-1997; 97EP-0122417.
 XX
 PR 18-DEC-1996; 96JP-0338673.
 XX
 PA (SUMO) SUMITOMO CHEM CO LTD.
 XX
 PI Oeda K, Wantanabe E;
 XX
 DR WPI: 1998-324670/29.
 DR N-PSDB: AAY40800.
 XX
 PT New nucleic acid molecule encoding plant raffinose synthetase -
 PT capable of producing raffinose, used as food additives with
 PT beneficial effects on gastrointestinal flora
 XX
 PS Claim 1: Page 26-29; 44pp; English.
 CC
 CC This sequence represents the broad bean raffinose synthetase of the
 CC invention. The raffinose synthetase is capable of producing raffinose by
 CC combining a D-galactosyl group through an alpha (1-6) bond with a
 CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
 CC residue in a sucrose molecule. The DNA can be used to modify metabolism
 CC of a host organism by introducing into the host organism or its cell so
 CC that the content of the raffinose family oligosaccharides in the host
 CC organism or cell is changed. Raffinose oligosaccharides are useful as
 CC food additives with beneficial effects on the gastrointestinal flora.
 XX
 SQ Sequence 799 AA:

Query Match 81.8%; Score 148; DB 19; Length 799;
 Best Local Similarity 80.0%; Pred. No. 8,1e-13;
 Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 FGMCTWDAFYLVHPGVIEGVRLVDGCG 30
 Db 238 FGMCTWDAFYLVHPGVIEGVRLVDGCG 267

RESULT 7
 AAY30142
 ID AAY30142 standard; Protein; 799 AA.
 AC AAY30142;
 XX
 DT 26-OCT-1999 (first entry)
 XX
 DE Amino acid sequence of a raffinose synthase protein.
 XX
 KM Raffinose synthase; plant; broad bean; sucrose; raffinose.
 XX
 OS Vicia faba.

DT 17-AUG-2001 (first entry)
XX
DE Soybean protein: SEQ ID 1.
XX
KW Mutant; mutain; raffinose synthase; raffinose oligosaccharide reduction;
KW plant; soybean.
XX
OS Glycine max.
XX
PN JP2001078783-A.
XX
PD 27-MAR-2001.
XX
XX
PF 03-JUL-2000; 2000JP-0200571.
XX
PR 09-JUL-1999; 99JP-0196036.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
XX
DR WPI; 2001-313373/33.
DR N-PSDB; AAH27438.
XX
PT Novel mutant protein of raffinose synthase is useful for reducing the
PT raffinose oligosaccharide content in a plant body -
XX
XX
PS Disclosure; Page 18-20; 30pp; Japanese.
XX
CC The present invention relates to a mutant protein of raffinose synthase
CC in which at least one aromatic amino acid present at the position of
CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
CC mutant protein can be used for reducing the raffinose oligosaccharide
CC content in a plant body. The present protein from soybean, was used in
CC the present invention.
XX
SQ Sequence 780 AA:

Query Match 80.7%; Score 146; DB 22; Length 780;
Best Local Similarity 80.0%; Pred. No. 1.5e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 30
DB 220 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 249

RESULT 11
AAW57887
ID AAW57887 standard; Protein: 781 AA.
XX
AC AAW57887;
XX
DT 23-SEP-1998 (first entry)
XX
DE Soybean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; soybean.
XX
OS Glycine max.
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.
XX
XX
PF 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Wantanabe E;
XX
DR WPI; 1998-324670/29.

DR N-PSDB; AAV40801.
XX
XX New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
XX
PS Claim 1: Page 31-34; 44pp; English.
XX
CC This sequence represents the soybean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
SQ Sequence 781 AA:

Query Match 80.7%; Score 146; DB 19; Length 781;
Best Local Similarity 80.0%; Pred. No. 1.5e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 30
DB 220 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 249

RESULT 12
AAV30143
ID AAV30143 standard; Protein: 781 AA.
XX
AC AAV30143;
XX
DT 26-OCT-1999 (first entry)
XX
DE Amino acid sequence of a raffinose synthase protein.
XX
KW Raffinose synthase; plant; sucrose; raffinose.
XX
OS Glycine max.
XX
PN JP11215984-A.
XX
PD 10-AUG-1999.
XX
PE 12-DEC-1997; 97JP-0342899.
XX
PR 28-NOV-1997; 97JP-0329006.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1999-511112/43.
DR N-PSDB; AAZ10002.
XX
PT New raffinose synthase gene - is prepared from a plant material
XX
PS Claim 8; Page 25-27; 40pp; Japanese.
XX
CC The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
SQ Sequence 781 AA:

Query Match 80.7%; Score 146; DB 20; Length 781;
Best Local Similarity 80.0%; Pred. No. 1.5e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 30

Db 220 FGWCTWDAFYLVKHPGVEGVKGLVEGCG 249

RESULT 13

AA849400
ID AAB49400 standard; Protein; 781 AA.

XX AAB49400;

DT 07-MAR-2001 (first entry)

DE Soybean raffinose synthase.

KW Plant promoter; transgenic plant; desired property.

OS Glycine max.

PN EP1048733-A2.

PD 02-NOV-2000.

PF 27-APR-2000; 2000EP-0108962.

PR 30-APR-1999; 99JP-0124527.

PR 01-SEP-1999; 99JP-0247211.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI Ishige F, Watanabe E, Oeda K;

DR WPI: 2001-104537/12.

DR N-PSDB; AAC89523.

PT New soybean plant promoters useful for generating transgenic plants

PS with desired properties -

CC Example 6; Page 24-27; 36pp; English.

CC The present invention provides novel plant promoters which can be used in

CC the production of transgenic plants which express genes with desired

CC properties.

CC Sequence 781 AA;

SO Query Match 80.7%; Score 146; DB 22; Length 781;

Best Local Similarity 80.0%; Pred. No. 1.5e-12;

Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 FGWCTWDAFYLVKHPGVEGVKGLVEGCG 30

DB 220 FGWCTWDAFYLVKHPGVEGVKGLVEGCG 249

RESULT 14

AA32075

ID AAY32075 standard; Protein; 572 AA.

XX AAY32075;

DT 17-JAN-2000 (first entry)

DE Rapeseed raffinose synthase.

KW Rapeseed raffinose synthase; rapeseed; transgenic plant.

OS Brassica napus.

PN Key Location/Qualifiers

FT Misc-difference 129 /note= "encoded by GGY"

FT Misc-difference 132 /note= "encoded by GGY"

FT Misc-difference 143 /note= "encoded by CCS"

FT Misc-difference 144 /note= "encoded by PCR"

FT Misc-difference 148 /note= "encoded by CCR"

FT Misc-difference 148 /note= "encoded by CCR"

PN EP953643-A2.

PD 03-NOV-1999.

PF 27-APR-1999; 99EP-0107430.

PR 30-APR-1998; 98JP-0120550.

PR 30-APR-1998; 98JP-0120551.

PR 04-DEC-1998; 98JP-0345590.

PR 10-DEC-1998; 98JP-0351246.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI Watanabe E, Oeda K;

DR WPI: 1999-593144/51.

DR N-PSDB; AA220210.

PT New sense and antisense genes, useful for altering the level of

PT raffinose in food plants -

PS Claim 27; Page 36-38; 55pp; English.

CC This sequence represents rapeseed raffinose synthase, a protein

CC that can bind a D-galactosyl group through an alpha(1-6) bond to the

CC hydroxy group attached to the carbon atom at the 6-position of the

CC D-glucose residue in a sucrose molecule to form raffinose. CDNA

CC (see AA220210) encoding the enzyme was isolated from rapeseed cv.

CC Westar leaf cDNA by PCR. Probes or primers generated from plant

CC raffinose synthase genes (see AA220210) may be used to obtain

CC other raffinose synthase genes by labeled detection or amplification

CC (claimed). These genes may be used to control the levels of

CC raffinose produced in plants. Antisense genes can be used to knock

CC out existing gene activity, and sense genes to increase the level

CC of gene activity. The resulting transgenic plants may be used as a

CC food source to alter the growing conditions for gut enterobacteria,

CC providing general health advantages.

CC Sequence 572 AA;

SO Query Match 79.6%; Score 144; DB 20; Length 572;

Best Local Similarity 80.0%; Pred. No. 2.1e-12;

Matches 24; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 FGWCTWDAFYLVKHPGVEGVKGLVEGCG 30

DB 13 FGWCTWDAFYLVKHPGVEGVKGLVEGCG 42

RESULT 15

AA32074

ID AAY32074 standard; Protein; 777 AA.

XX AAY32074;

DT 17-JAN-2000 (first entry)

DE Mustard raffinose synthase.

KW Mustard raffinose synthase; mustard; transgenic plant.

OS Brassica juncea.

PN Key Location/Qualifiers

FT Misc-difference 210 /note= "encoded by ACR"

FT Misc-difference 210 /note= "encoded by ACR"

XX EP953643-A2.
 PN
 XX
 PD 03-NOV-1999.
 XX
 XX 27-APR-1999; 99EP-0107430.
 PF
 XX 30-APR-1998; 98JP-0120550.
 PR 30-APR-1998; 98JP-0120551.
 PR 04-DEC-1998; 98JP-0345590.
 PR 10-DEC-1998; 98JP-0351246.
 XX
 PA (SUMO) SUMITOMO CHEM CO LTD.
 XX
 PI Watanabe E, Oeda K;
 XX
 XX WPI: 1999-593144/51.
 DR N-PSDB; AA220209.
 PT
 XX New sense and antisense genes, useful for altering the level of
 PT raffinose in food plants -
 XX
 PS Claim 26; Page 29-31; 55pp; English.
 XX
 CC This sequence represents mustard raffinose synthase, a protein
 CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
 CC hydroxy group attached to the carbon atom at the 6-position of the
 CC D-glucose residue in a sucrose molecule to form raffinose. CDNA
 CC (see AA220209) encoding the enzyme was isolated from mustard
 CC (Brassica juncea) leaf cDNA by PCR. Probes or primers generated from
 CC plant raffinose synthase genes (see AA220207-10) may be used to obtain
 CC other raffinose synthase genes by labeled detection or amplification
 CC (claimed). These genes may be used to control the levels of
 CC raffinose produced in plants. Antisense genes can be used to knock
 CC out existing gene activity, and sense genes to increase the level
 CC of gene activity. The resulting transgenic plants may be used as a
 CC food source to alter the growing conditions for gut enterobacteria,
 CC providing general health advantages.
 XX
 SO Sequence 777 AA;

Query Match 79.6%; Score 144; DB 20; Length 777;
 Best Local Similarity 80.0%; Pred. No. 3e-12;
 Matches 24; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 OY 1 FGMCTDAFYLTVPQGVYEGVRLVDGCG 30
 |||
 Db 218 FGMCTDAFYLTVPQGVYEGVRLVDGCG 247

Search completed: March 27, 2003, 10:03:41
 Job time : 22.2195 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 8.59756 Seconds
(without alignments)
335.448 Million cell updates/sec

Title: US-09-675-208-1
Perfect score: 181
Sequence: 1 FGWCTWDAFYLTWHPQGVIEGVRLVDGCG 30

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	111	61.3	765	2 S45033	probable imbibition
2	108	59.7	357	2 T09530	probable seed inh1
3	108	59.7	757	2 S27762	S1p1 protein - bar
4	108	59.7	773	2 T46188	Imbibition protein
5	106	58.6	807	2 C95025	hypothetical prote
6	82	45.3	1170	2 C96599	protein F20N2.14 l
7	72	39.8	371	2 T01717	hypothetical prote
8	64.5	35.6	374	2 E82168	probable alpha-1,6
9	60	33.1	137	2 I46711	fibroblast growth
10	60	33.1	146	1 S00185	basic fibroblast g
11	60	33.1	154	2 A31674	basic fibroblast g
12	60	33.1	154	2 C37360	basic fibroblast g
13	60	33.1	157	1 GKB08	basic fibroblast g
14	60	33.1	210	2 A23398	basic fibroblast g
15	60	33.1	451	2 T24018	hypothetical prote
16	59	32.6	164	2 S31622	basic fibroblast g
17	57.5	31.8	649	2 D90496	hypothetical prote
18	56.5	31.2	373	2 A10440	probable aliphatic
19	53.5	29.6	319	2 C90756	hypothetical prote
20	53.5	29.6	333	2 A85620	hypothetical prote
21	53.5	29.6	333	2 C64833	probable nitrate t
22	53	29.3	189	2 A48834	basic fibroblast g
23	53	29.3	557	2 E83744	hypothetical prote
24	52.5	29.0	265	2 E82964	membrane subunit o
25	52	28.7	237	2 B86420	unknown protein, 1
26	52	28.7	619	2 D86509	S/T protein kinase
27	52	28.7	619	2 D81556	conserved hypothet
28	52	28.7	619	2 A72114	s/t protein kinase
29	52	28.7	1462	1 B36182	protein-tyrosine-p

30	51.5	28.5	353	2 AE2067	histidinol phosphat
31	51	28.2	106	1 ERAD31	early E3a 12.1K pr
32	51	28.2	106	2 S52799	hypothetical prote
33	51	28.2	106	2 S52809	hypothetical prote
34	51	28.2	225	2 AD0798	probable phosphata
35	51	28.2	1039	2 F71427	hypothetical prote
36	50	27.6	284	2 AD0533	conserved hypothet
37	50	27.6	381	2 T23250	Ampe protein (limpo
38	50	27.6	996	2 F98092	hypothetical prote
39	50	27.6	996	2 F98092	cytM protein, cyto
40	50	27.6	996	2 A85228	bacteriocin format
41	49.5	27.3	243	2 AC3600	cellulase (EC 3.2.
42	49.5	27.3	310	2 S39592	prolyl aminopeptid
43	49.5	27.3	720	2 AF0242	probable exported
44	49	27.1	79	2 S55456	GAM1 protein - Pla
45	49	27.1	191	1 H64744	yaed protein - Esc

ALIGNMENTS

RESULT 1

S45033
Probable imbibition protein - wild cabbage
C:Species: Brassica oleracea (wild cabbage)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S45033
R:Fujikura, Y.; Karsen, C.K.
Submitted to the EMBL Data Library, May 1994
A:Description: Cauliflower cDNA encoding a putative imbibition protein.
A:Reference number: S45033
A:Accession: S45033
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-765 <FUJ>
A:Cross-references: EMBL:X79330; NID:q488786; PID:q488787

Query Match 61.3%; Score 111; DB 2; Length 765;
Best Local Similarity 65.5%; Pred. No. 2.3e-07;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTWHPQGVIEGVRLVDGCG 29

DB 198 FGWCTWDAFYLTWHPQGVIEGVRLVDGCG 226

RESULT 2

T09530
Probable seed imbibition protein - chickpea (fragment)
C:Species: Cicer arietinum (chickpea, garbanzo)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T09530
R:Carver, E.
Submitted to the EMBL Data Library, February 1996
A:Reference number: Z16718
A:Accession: T09530
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-357 <CEP>
A:Cross-references: EMBL:X95875; NID:q1212811
A:Experimental source: germinating seed
C:Genetics:

QY 1 FGWCTWDAFYLTWHPQGVIEGVRLVDGCG 29

DB 31 FGWCTWDAFYLTWHPQGVIEGVRLVDGCG 59

Query Match 59.7%; Score 108; DB 2; Length 357;
Best Local Similarity 62.1%; Pred. No. 2.6e-07;
Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Db 111 FGCTWDAFYRKV 123

RESULT 8

E82168

Probable alpha-1,6-galactosidase VC1690 [Imported] - Vibrio cholerae (strain N16961 sero

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: E82168

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: E82168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-374 <HEI>

A:Cross-references: GB:AE004247; GB:AE003852; NID:99562204; PIDN:AAE94840.1; GSPDB:GN001

C:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC1690

A:Map position: 1

Query Match 35.6%; Score 64.5; DB 2; Length 374;

Best Local Similarity 38.5%; Pred. No. 0.32;

Matches 15; Conservative 4; Mismatches 9; Indels 11; Gaps 2;

Oy 2 GCCTWDAFYLVHPOGVIEGV-----RH-----LVDCG 29

Db 197 GWCWYAYAEVTEODIKENVAIIAERHPELEWVLLDDG 235

RESULT 9

I46711

Fibroblast growth factor - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 16-Jul-1999

C:Accession: I46711

R:Winkler, J.A.; Erlengel, R.; Alberts, G.F.; Janat, M.F.; Lian, G.

Ann. N. Y. Acad. Sci. 143, 518-527, 1993

A:Title: Elevated expression of basic fibroblast growth factor in an immortalized rabbit

A:Reference number: I46711; MUID:93343209; PMID:8342599

A:Accession: I46711

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-137 <WIN>

A:Cross-references: GB:LI2034; NID:9165014; PIDN:AAA31248.1; PID:9165015

C:Superfamily: fibroblast growth factor

Query Match 33.1%; Score 60; DB 2; Length 137;

Best Local Similarity 40.0%; Pred. No. 0.47;

Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Oy 3 WCTWDAFYLVHPOGVIEGVRLHVD 27

Db 24 YCKNGCFRLRHDPGRVDGVRKSD 48

RESULT 10

S00185

basic fibroblast growth factor - sheep

N:Alternate names: prostathionin

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S00185

R:Simpson, R.J.; Moritz, R.L.; Lloyd, C.J.; Fabri, L.J.; Nice, E.C.; Rudira, M.R.; Burge

FEBS Lett. 224, 128-132, 1987

A:Title: Primary structure of ovine pituitary basic fibroblast growth factor.

A:Reference number: S00185; MUID:88055577; PMID:3678486

A:Accession: S00185

A:Molecule type: protein

A:Residues: 1-146 <SIM>
C:Superfamily: fibroblast growth factor
C:Keywords: growth factor; heparin binding; mitogen
F:18-22/Region: heparin binding #status predicted
F:107-110/Region: heparin binding #status predicted

Query Match 33.1%; Score 60; DB 1; Length 146;

Best Local Similarity 40.0%; Pred. No. 0.5;

Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Oy 3 WCTWDAFYLVHPOGVIEGVRLHVD 27

Db 24 YCKNGCFRLRHDPGRVDGVRKSD 48

RESULT 11

A31674

basic fibroblast growth factor precursor - rat

N:Alternate names: bFGF

C:Species: Rattus norvegicus (Norway rat)

C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 16-Jul-1999

C:Accession: A31674; S00876; S24309

R:Shimasaki, S.; Emoto, N.; Koba, A.; Mercado, M.; Shibata, F.; Cooksey, K.; Baird, A

Biochem. Biophys. Res. Commun. 157, 236-263, 1988

A:Title: Complementary DNA cloning and sequencing of rat ovarian basic fibroblast gro

A:Reference number: A31674; MUID:89061721; PMID:3196337

A:Accession: A31674

A:Molecule type: mRNA

A:Residues: 1-154 <SHI>

A:Cross-references: GB:M22427; NID:9204285; PIDN:AAA41210.1; PID:9204286

R:Kurokawa, T.; Seno, M.; Igarashi, K.

Nucleic Acids Res. 16, 5201, 1988

A:Title: Nucleotide sequence of rat basic fibroblast growth factor cDNA.

A:Reference number: S00876; MUID:88262516; PMID:3387229

A:Accession: S00876

A:Molecule type: mRNA

A:Residues: 1-154 <KUR>

A:Cross-references: EMBL:X07285; NID:956203; PIDN:CAA30265.1; PID:956204

R:El-Husseini, A.E.D.; Paterson, J.A.; Myal, Y.; Shiu, R.P.C.

Biochim. Biophys. Acta 1131, 314-316, 1992

A:Title: PCR detection of the rat brain basic fibroblast growth factor (bFGF) mRNA co

A:Reference number: S24309; MUID:92329546; PMID:1378302

A:Accession: S24309

A:Status: preliminary; translation not shown

A:Molecule type: mRNA

A:Cross-references: EMBL:X61697; NID:956143; PIDN:CAA43863.1; PID:956144

C:Superfamily: fibroblast growth factor

C:Keywords: growth factor

F:1-9/Domain: signal sequence #status predicted <SIG>

F:10-154/Product: basic fibroblast growth factor #status predicted <MAT>

Query Match 33.1%; Score 60; DB 2; Length 154;

Best Local Similarity 40.0%; Pred. No. 0.53;

Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Oy 3 WCTWDAFYLVHPOGVIEGVRLHVD 27

Db 32 YCKNGCFRLRHDPGRVDGVRKSD 56

RESULT 12

C37360

basic fibroblast growth factor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 16-Jul-1999

C:Accession: C37360

R:Hebert, J.M.; Basilico, C.; Goldfarb, M.; Haub, O.; Martin, G.R.

Dev. Biol. 138, 454-463, 1990

A:Title: Isolation of cDNAs encoding four mouse FGF family members and characterization

A:Reference number: A37360; MUID:90201563; PMID:2318343

A:Accession: C37360

A:Status: preliminary

[illegible]

A:Contents: amino end of this form was blocked; the peptide composition matched what
A>Note: the amino end of this form was blocked; the peptide composition matched what
R.Gospodarowicz, D.; Baird, A.; Cheng, J.; Lai, G.M.; Esch, F.; Bohlen, P.
Endocrinology 118, 82-90, 1986
A>Title: Isolation of fibroblast growth factor from bovine adrenal gland: physicochemical
A:Reference number: A61094; MUID:86081530; PMID:3940857
A:Accession: A61094
A:Molecule type: protein
A:Residues: 12-25,27-35,'X',37-40 <GOS>
A:Experimental source: adrenal gland
R:Esch, F.; Baird, A.; Ling, N.; Ueno, N.; Hill, F.; Denoroy, L.; Klepper, R.; Gospod
Proc. Natl. Acad. Sci. U.S.A. 82, 6507-6511, 1985
A>Title: Primary structure of bovine pituitary basic fibroblast growth factor (bFGF) a
A:Reference number: A01386; MUID:86016731; PMID:3863109
A:Accession: A01386
A:Molecule type: protein
A:Residues: 12-15' <ESC>
A:Experimental source: pituitary gland
R:Baird, A.; Esch, F.; Boehlen, P.; Ling, N.; Gospodarowicz, D.
Regul. Pept. 12, 201-213, 1985
A>Title: Isolation and partial characterization of an endothelial cell growth factor
A:Reference number: A60316; MUID:86095426; PMID:4081126
A:Accession: A60316
A:Molecule type: protein
A:Residues: 27-35,'X',37-43 <BAI>
A:Experimental source: kidney
R:Bohlen, P.; Baird, A.; Esch, F.; Ling, N.; Gospodarowicz, D.
Proc. Natl. Acad. Sci. U.S.A. 81, 5364-5368, 1984
A>Title: Isolation and partial molecular characterization of pituitary fibroblast gro
A:Reference number: A22054; MUID:84298139; PMID:6591194
A:Accession: A22054
A:Molecule type: protein
A:Residues: 12-26 <BOH>
C:Comment: The acidic and basic fibroblast growth factors are the major endothelial-c
ell types in vitro (although bFGF is 30-100 times more potent than aFGF in stimulat
in C:Comment: This protein binds heparin more strongly than does aFGF.
C:Superfamily: fibroblast growth factor
C:Keywords: alternative splicing; angiogenesis; blocked amino end; growth factor; hepa
F:1-157/Product: basic fibroblast growth factor, uterine form #status predicted <MA1
F:4-157/Product: basic fibroblast growth factor, pituitary gamma form #status experim
F:12-157/Product: basic fibroblast growth factor, pituitary alpha form #status experi
F:16-157/Product: basic fibroblast growth factor, pituitary short form #status predic
F:27-157/Product: basic fibroblast growth factor, hepatic form #status experimental <M
F:29-33,118-121/Region: heparin binding #status predicted
F:4/Modified site: Blocked amino end (Ala) (In mature form pituitary gamma) (probably

Query Match 33.1%; Score 60; DB 1; Length 157;
Best Local Similarity 40.0%; Pred. No. 0.54;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 3 WCTWDARLYVHPGVIEGVRLVD 27
: : | : | : | : | : | : | :
ob 35 YCKNGFFLRIRHPGRVGDVEKSD 59

RESULT 14
A32398
basic fibroblast growth factor precursor, 22.5k form - human
N:Alternate names: bFGF; fibroblast growth factor 2; prostatic growth factor; prostat
N:Contains: basic fibroblast growth factor, 18k form
C:Species: Homo sapiens (man)
C>Date: 31-Jul-1989 #sequence, revision 31-Dec-1993 #text, change 21-Jul-2000
C:Accession: A32398; A61537; A26642; B32878; S00257; A54116; B54316; A33624; A25824;
R:Prats, H.; Kagniad, M.; Prats, A.C.; Klagesorn, M.; Dellas, J.M.; Liauzun, P.; Chailo
Proc. Natl. Acad. Sci. U.S.A. 86, 1836-1840, 1989
A>Title: High molecular mass forms of basic fibroblast growth factor are initiated by
A:Reference number: A32398; MUID:89184522; PMID:2533817
A:Accession: A32398
A:Molecule type: PRNA
A:Residues: 1-210 <PRA>
A:Cross-references: GB:U04513; MID:g183083; PIDN:AAA52531.1; PID:g459811
Shibata, F.; Baird, A.; Florkiewicz, R.Z.

Growth Factors 4, 277-287, 1991

A:Title: Functional characterization of the human basic fibroblast growth factor gene
A:Reference number: A61537; MUID:92110035; PMID:11764264

A:Accession: A61537

A:Molecule type: DNA

A:Residues: 1-114 <SH1>

A>Note: authors translated the codon GGA for residue 47 as Ala

R:Kurakawa, T.; Sasada, R.; Iwane, M.; Igarashi, K.

FEBS Lett. 213, 189-194, 1987

A:Title: Cloning and expression of cDNA encoding human basic fibroblast growth factor.

A:Reference number: A26642; MUID:87162468; PMID:2435575

A:Accession: A26642

A:Molecule type: mRNA

A:Residues: 56-210 <KUR>

A:Cross-references: GB:427968; NID:g182562; PIDN:AAA5448.1; PID:g182563

R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergle, A.; Fildes, J.C.

Cold Spring Harb. Symp. Quant. Biol. 51, 657-668, 1986

A:Title: Human basic fibroblast growth factor: nucleotide sequence, genomic organization

A:Reference number: A90924; MUID:87210066; PMID:3472745

A:Accession: B32878

A:Molecule type: mRNA

A:Residues: 56-210 <ABR>

A>Note: the authors translated the codon GAA for residue 108 as Gly

R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergle, A.; Friedman, J.; Gospodarowicz, D.; E

EMBO J. 5, 2523-2528, 1986

A:Title: Human basic fibroblast growth factor: nucleotide sequence and genomic organization

A:Reference number: S00297; MUID:87053817; PMID:3780670

A:Accession: S00297

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-155 <AB2>

A>Note: the authors translated the codon GAA for residue 108 as Gly

R:Shimoyama, Y.; Gotch, M.; Ino, Y.; Sakamoto, M.; Kato, K.; Hirohashi, S.

Jpn. J. Cancer Res. 82, 1263-1270, 1991

A:Title: Characterization of high-molecular-mass forms of basic fibroblast growth factor

clonogenesis.

A:Reference number: A54316; MUID:92091228; PMID:1721615

A:Accession: A54316

A:Molecule type: protein

A:Residues: 'XX', '86-88', 'X', '90-91', 'X', '93-95' <SH3>

A:Experimental source: C-1121 hepatocellular carcinoma cell line

A>Note: sequence extracted from NCBI backbone (NCBIP:71595)

A:Accession: B54316

A:Molecule type: protein

A:Residues: 'XXX', '19', 'X', '21-29' <SH2>

A>Note: sequence extracted from NCBI backbone (NCBIP:71594)

R:Fejge, J.J.; Bradley, J.D.; Fryburg, K.; Farris, J.; Cousins, L.C.; Barr, P.J.; Baird,

J. Cell Biol. 109, 3105-3111, 1989

A:Title: Differential effects of heparin, fibronectin, and laminin on the phosphorylation

A:Reference number: A33624; MUID:90078343; PMID:2592418

A:Accession: A33624

A>Status: preliminary

A:Molecule type: protein

A:Residues: 57-210 <FE1>

R:Story, M.T.; Esch, F.; Shimazaki, S.; Sasse, J.; Jacobs, S.C.; Lawson, R.K.

Biochem. Biophys. Res. Commun. 142, 702-709, 1987

A:Title: Amino-terminal sequence of a large form of basic fibroblast growth factor isolated

A:Reference number: A25824; MUID:87156686; PMID:2435284

A:Accession: A25824

A:Molecule type: protein

A:Residues: 57-77 <STO>

A:Experimental source: prostate

R:Gimenez-Gallo, G.; Conn, G.; Hatcher, V.B.; Thomas, K.A.

Biochem. Biophys. Res. Commun. 135, 541-548, 1986

A:Title: Human brain-derived acidic and basic fibroblast growth factors: amino terminal

A:Reference number: A90122; MUID:86186784; PMID:3964255

A:Accession: B24243

A:Molecule type: protein

A:Residues: 65-102, 'X', '104-105' <GIM>

A:Experimental source: brain

R:Gautschi, P.; Frazer-Schroder, M.; Bohlen, P.

FEBS Lett. 204, 203-207, 1986

A:Title: Partial molecular characterization of endothelial cell mitogens from human brain

```

A:Reference number: A91364; MUID:86275260; PMID:3732516
A:Molecule type: protein
A:Residues: 65-88,'X','90-98','X',100 <GAD>
R:Sommer, A.; Brewer, M.T.; Thompson, R.C.; Moscatelli, D.; Presta, M.; Rifkin, D.B.
Biochem. Biophys. Res. Commun. 144, 543-550, 1987
A>Title: A form of human basic fibroblast growth factor with an extended amino termin
A:Reference number: S42424; MUID:8721328; PMID:3579950
A:Accession: S42424
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 54-210 <SOM>
A:Cross-references: EMBL:M17599; NID:g183086; PIDN:AAS52534.1; PID:g183087
R:Patellano, M.W.; Horlick, R.A.; Springer, B.A.; Van Dyk, D.E.; Tobey, T.; Wetmore
Biochemistry 33, 10229-10246, 1994
A>Title: Multivalent ligand-receptor binding interactions in the fibroblast growth fa
A:Reference number: A55784; MUID:94347757; PMID:7520751
A:Accession: B55784
A:Molecule type: protein
A:Residues: 54-71 <PAN>
R:Watson, R.; Anthony, F.; Pickett, M.; Lambden, P.; Masson, G.M.; Thomas, E.J.
Biochem. Biophys. Res. Commun. 187, 1227-1231, 1992
A>Title: Reverse transcription with nested polymerase chain reaction shows expresio
lients.
A:Reference number: I52267; MUID:93038590; PMID:1417798
A:Accession: I52267
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 95-182 <RES>
A:Cross-references: GB:S47380; NID:g256535; PIDN:AMD1853.1; PID:g4261553
R:Experty, V.; Buglier, B.; Amalric, F.; Prome, J.C.; Prats, H.
FEBS Lett. 349, 23-28, 1994
A>Title: Purification and characterization of the 210-amino acid recombinant basic fi
A:Reference number: S46253; MUID:94320639; PMID:8045296
A:Accession: S46253
A:Molecule type: protein
A:Residues: 39-53;65-88 <PAT>
A>Note: recombinant gene expressed in Escherichia coli
C:Genetics:
A:Gene: GDB:FGF2; FGFB
A:Cross-references: GDB:I19910; OMIM:134920
A:Map position: 4q25-4q27
A:Start codon: CCG
C:Keywords: alternative initiators; angiogenesis; growth factor; heparin binding; mit
E:1-210/Product: basic fibroblast growth factor, 22.5K form #status predicted <M2>
E:65-210/Product: basic fibroblast growth factor, 18k form #status predicted <M2>
E:82-86/Region: heparin binding #status predicted
E:171-174/Region: heparin binding #status predicted
Query Match 33.1%; Score 60; DB 2; Length 210;
Best Local Similarity 40.0%; Pred. No. 0.73;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
OY 3 WCTWDATFLVTHPGVIEGVRLHVD 27
:| :| :| :| :| :|
DB 88 YCKNGCFELRIHPDGRVDGVREKSD 112
RESULT 15
T24018
hypothetical protein R07B7.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T24018
R:Harris, B.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19830
A:Accession: T24018
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-451 <WIL>

```

A:Cross-references: EMBL:Z75955; PIDN:CAB00120.1; GSPDB:GN00023; CESP:R07B7.11
 A:Experimental source: clone R07B7
 C:Genetics:
 A:Gene: CESP:R07B7.11
 A:Map position: 5
 A:Introns: 61/1,107/3; 317/3; 351/3
 C:Superfamily: alpha-galactosidase

Query Match 33.1%; Score 60; DB 2; Length 451;
 Best Local Similarity 36.8%; Pred. No. 1.7;
 Matches 14; Conservative 3; Mismatches 11; Indels 10; Gaps 2;

QY 2 GWCCTDPAFLTV---HPQGV-----EGVRHLVDG 29
 ||:|||||:|||||:|||||
 Db 27 GWMSTAFYCEIDCVKHPGTCINEQLYKMDADQIVSGG 64

Search completed: March 27, 2003, 10:06:47
 Job time : 10.5976 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 4.93902 Seconds

(Without alignments)
251.930 Million cell updates/sec

Title: US-09-675-208-1

Percent score: 181

Sequence: 1 FGWCTMDAFYLTWHPQGVIEGVRLVDGCG 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	33.1	137	1	FGF2_RABIT
2	60	33.1	154	1	FGF2_MOUSE
3	60	33.1	154	1	FGF2_RAT
4	60	33.1	155	1	FGF2_BOVIN
5	60	33.1	155	1	FGF2_HUMAN
6	60	33.1	155	1	FGF2_SHEEP
7	59	32.6	156	1	FGF2_MONDO
8	53.5	29.6	319	1	SSUA_ECOLI
9	53	29.3	158	1	FGF2_CHICK
10	52	28.7	1462	1	PRP6_DROME
11	51.5	28.5	353	1	H181_AMASP
12	51	28.2	106	1	E312_ADE03
13	49.5	27.3	310	1	PIP_NEIGO
14	49	27.1	191	1	VAED_ECOLI
15	49	27.1	227	1	LIPIB_STY3
16	49	27.1	390	1	DXR_FUSNN
17	48.5	26.8	302	1	YIAL_PSESS
18	48.5	26.8	515	1	GLCM_MOUSE
19	48.5	26.8	536	1	GLCM_HUMAN
20	48	26.5	207	1	RGFG_HUMAN
21	48	26.5	207	1	RGFG_RAT
22	48	26.5	366	1	CHSB_IPOFE
23	48	26.5	384	1	DXR_CLOPE
24	48	26.5	388	1	DXR_ZYMO
25	48	26.5	468	1	LACG_STRMU
26	48	26.5	470	1	LACG_STANU
27	47.5	26.2	369	1	MAGA_HUMAN
28	47.5	26.2	500	1	C72X_ARATH
29	47.5	26.2	500	1	C72Y_ARATH
30	47	26.0	247	1	FGFE_HUMAN
31	47	26.0	247	1	FGFE_MOUSE
32	47	26.0	341	1	SP12_COMPX
33	47	26.0	361	1	U204_ARATH

34	47	26.0	382	1	DXR_THETN	08a28 thermomae
35	47	26.0	406	1	DXR_MYCLE	09cb3 mycobacteri
36	47	26.0	413	1	DXR_MYCTU	Q10798 mycobacteri
37	47	26.0	548	1	HEML_YEAST	P09950 saccharomyc
38	47	26.0	686	1	VILL_HUMAN	O15195 homo sapien
39	46.5	25.7	299	1	RFB1_SALT	P22716 salmonella
40	46.5	25.7	498	1	NDDO_ALCX	P94212 alcaligenes
41	46.5	25.7	521	1	OPN4_MOUSE	Q9qxx9 mus musculu
42	46.5	25.7	582	1	HEMO_OPSTA	P43090 opsanus lau
43	46	25.4	132	1	VEA_HPVA	P25483 human papil
44	46	25.4	222	1	SP1B_VACCC	P20842 vaccinia vi
45	46	25.4	352	1	CHSC_IPOPU	P48399 ipomoea pur

ALIGNMENTS

RESULT 1
FGF2_RABIT STANDARD: PRT: 137 AA.

AC P48799;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 (HGF-2) (Basic fibroblast growth factor) (BFGF) (Prostathelin) (Fragment).
GN FGF2.

OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]

RP STRAIN=New Zealand white; TISSUE=Smooth muscle;
RC MEDLINE=9343209; PubMed=8342599;
RX Winkles J.A., Friesel R., Alberts G.F., Janat M.F., Liu G.;
RT "Elevated expression of basic fibroblast growth factor in an
Am. J. Pathol. 143:518-527(1993)."
RL - FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS
CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN
CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND
CC CONCENTRATION OF THESE 2 GROWTH FACTORS.

CC - SUBUNIT: MONOMER.
CC - MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES
CC AEGF.

CC - SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

CC EMBL: L12034; AAA31248.1; -
CC HSSP: P09038; 1BPF.
CC InterPro: IPR002209; HB/F-growthfact.
DR Pfam: PF00167; FGF, 1.
DR ProDom: PD000831; HB/F-growthfact; 1.
DR SMART: SM00442; FGF, 1.
DR PROSITE: PS00247; HBGF_FGF, 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT BINDING 18 22 HEPARIN (POTENTIAL).
FT NON_TER 107 110 HEPARIN (POTENTIAL).
FT SEQUENCE 137 AA: 15418 MW: 0D9EE457B88BC51 CRC64;
SQ

Query Match 33.1%; Score 60; DB 1; Length 137;
Best Local Similarity 40.0%; Pred. NO. 0.18; Mismatches 5; Indels 0; Gaps 0;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

DR PIR: S00876; S00876.
 DR PIR: A31674; A31674.
 DR HSSP: P09038; 1BPF.
 DR InterPro: IPR002209; HB/F-growthfact.
 DR InterPro: IPR002348; IL1_HBGF.
 DR Pfam: PF00167; FGF; 1.
 DR PRINTS: PR00262; IL1HBGF.
 DR ProDom: PD000831; HB/F-growthfact; 1.
 DR SMART: SM00442; FGF; 1.
 DR PROSITE: PS00247; HBGF_FGF; 1.
 KM Growth factor; Mitogen; Angiogenesis; Heparin-binding.
 FT PROPEP 1 9
 FT CHAIN 10 154 HEPARIN-BINDING GROWTH FACTOR 2.
 FT BINDING 26 30 HEPARIN (POTENTIAL).
 FT BINDING 115 118 HEPARIN (POTENTIAL).
 SQ SEQUENCE 154 AA: 17139 MW: 1A0F14FA23D8403 CRC64:
 Query Match 33.1%; Score 60; DB 1; Length 154;
 Best Local Similarity 40.0%; Pred. No. 0.21;
 Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
 QY 3 WCTMDAFYTVHPOGVIEGVRLVD 27
 DB 32 YCKNGCFELRHHPDGRVDSVREKSD 56
 RESULT 4
 FGF2_BOVIN STANDARD: PRT: 155 AA.
 ID FGF2_BOVIN STANDARD: PRT: 155 AA.
 AC P03969;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Heparin-binding growth factor 2 precursor (HBGF-2) (basic fibroblast growth factor) (aFGF) (Prostatropin) [contains: kidney-derived growth factor].
 DE FGF2 OR FGF-2.
 GN Bos taurus (Bovine).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; CC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 OX 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE-86261806; PubMed-2425435;
 RA Abraham J.A., Mergia A., Whang J.L., Tumolo A., Friedman J., Hjerlild K.A., Gospodarowicz D., Fiddes J.C.;
 RT "Nucleotide sequence of a bovine clone encoding the angiogenic protein, basic fibroblast growth factor.";
 RL Science 233:545-548(1986).
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87217066; PubMed-3472745;
 RA Abraham J.A., Whang J.L., Tumolo A., Mergia A., Fiddes J.C.;
 RT "Human basic fibroblast growth factor: nucleotide sequence, genomic organization, and expression in mammalian cells.";
 RL Cold Spring Harb. Symp. Quant. Biol. 51:657-668(1986).
 RN 13
 RP SEQUENCE OF 10-155.
 RX MEDLINE-86016731; PubMed-3863109;
 RA Esch F., Baird A., Ling N., Ueno N., Hili F., Denoroy L., Klepper R., Gospodarowicz D., Boehlen P., Guillemin R.;
 RT "Primary structure of bovine pituitary basic fibroblast growth factor (FGF) and comparison with the amino-terminal sequence of bovine brain acidic FGF.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:6507-6511(1985).
 RN 14
 RP SEQUENCE OF 1-9.
 RX MEDLINE-86295737; PubMed-3741423;
 RA Ueno N., Baird A., Esch F., Ling N., Guillemin R.;
 RT "Isolation of an amino terminal extended form of basic fibroblast growth factor.";
 RL Biochem. Biophys. Res. Commun. 138:580-588(1986).

RN 15
 RP SEQUENCE OF 25-41.
 RC TISSUE=Kidney;
 RX MEDLINE-86095426; PubMed-4081126;
 RA Baird A., Esch F., Boehlen P., Ling N., Gospodarowicz D.;
 RT "Isolation and partial characterization of an endothelial cell growth factor from the bovine kidney: homology with basic fibroblast growth factor.";
 RL Regul. Pept. 12:201-213(1985).
 RN 16
 RP SEQUENCE OF 21-40.
 RC TISSUE=Kidney;
 RX MEDLINE-87119165; PubMed-3809608;
 RA Ueno N., Baird A., Esch F., Shimasaki S., Ling N., Guillemin R.;
 RT "Purification and partial characterization of a mitogenic factor from bovine liver: structural homology with basic fibroblast growth factor.";
 RL Regul. Pept. 16:135-145(1986).
 RN 17
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE-91095983; PubMed-1702556;
 RA Zhu X., Komiyama H., Chirino A., Faham S., Fox G.M., Arakawa T., Hsu B.T., Rees D.C.;
 RT "Three-dimensional structures of acidic and basic fibroblast growth factors.";
 RL Science 251:90-93(1991).
 CC -!- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND CONCENTRATION OF THESE 2 GROWTH FACTORS.
 CC -!- SUBUNIT: MONOMER.
 CC -!- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES AFGF.
 CC -!- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: M13440; AAA0518.1; -;
 DR PIR: A24653; GKBOB.
 DR PIR: A24819; A24819.
 DR PIR: A32878; A32878.
 DR PDB: 1BAS; 3I-OCT-93.
 DR InterPro: IPR002209; HB/F-growthfact.
 DR InterPro: IPR002348; IL1_HBGF.
 DR Pfam: PF00167; FGF; 1.
 DR PRINTS: PR00262; IL1HBGF.
 DR ProDom: PD000831; HB/F-growthfact; 1.
 DR SMART: SM00442; FGF; 1.
 DR PROSITE: PS00247; HBGF_FGF; 1.
 KM Growth factor; Mitogen; Angiogenesis; Heparin-binding;
 RN 3D-structure.
 FT PROPEP 1 9
 FT CHAIN 10 155 HEPARIN-BINDING GROWTH FACTOR 2.
 FT SITE 25 155 KIDNEY-DERIVED GROWTH FACTOR.
 FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 88 90 CELL ATTACHMENT SITE (POTENTIAL).
 FT BINDING 27 31 HEPARIN (POTENTIAL).
 FT BINDING 116 119 HEPARIN (POTENTIAL).
 FT STRAND 30 34
 FT TURN 35 38
 FT STRAND 39 43
 FT TURN 44 46
 FT STRAND 47 52
 FT TURN 53 56
 FT HELIX 58 60
 FT STRAND 62 68
 FT TURN 69 70

RT spectroscopy.;

RL Biochemistry 35:13552-13561(1996).

CC -1- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS

CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN

CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND

CC CONCENTRATION OF THESE 2 GROWTH FACTORS.

CC -1- SUBUNIT: MONOMER.

CC -1- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES

CC AFGF.

CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>

CC or send an email to license@isb-sdb.ch).

CC -----

DR EMBL; M17599; AAA52534.1; ALT_INIT.

DR EMBL; X04431; CA28027.1; -.

DR EMBL; X04432; CA28028.1; -.

DR EMBL; X04433; CA28029.1; -.

DR EMBL; M27968; AAA52448.1; -.

DR EMBL; J04513; AAA52533.1; ALT_INIT.

DR PIR; A25824; A25824.

DR PIR; A26642; A26642.

DR PIR; B24243; B24243.

DR PIR; B24301; B24301.

DR PIR; B32878; B32878.

DR PIR; S00297; S00297.

DR PDB; 2EGF; 15-APR-92.

DR PDB; 4FGF; 15-JUL-93.

DR PDB; 1FGA; 15-JUL-93.

DR PDB; 1BFB; 03-APR-96.

DR PDB; 1BFC; 03-APR-96.

DR PDB; 1BFG; 16-JUN-97.

DR PDB; 1BFG; 31-JAN-94.

DR PDB; 2BFH; 30-APR-94.

DR PDB; 1BLA; 08-NOV-96.

DR PDB; 1BLD; 08-NOV-96.

DR Genem; HGNC:3676; FGF2.

DR MIM; 134920; -.

DR InterPro; IPR002209; HB/F-growthfact.

DR InterPro; IPR002348; ILI_HBGF.

DR Pfam; PF00167; FGF; 1.

DR PRINTS; PR00262; ILI_HBGF.

DR PRODom; PD000831; HB/F-growthfact; 1.

DR SMART; SM00442; FGF; 1.

DR PROSITE; PS00247; HBGF_FGF; 1.

DR Growth factor; Mitogen; Angiogenesis; Heparin-binding;

DR 3D-structure.

FT PROPEP 1 9

FT CHAIN 10 155

FT SITE 46 48

FT SITE 88 90

FT BINDING 27 31

FT BINDING 116 119

FT STRAND 30 34

FT TURN 35 38

FT STRAND 39 43

FT TURN 45 45

FT STRAND 49 52

FT TURN 55 56

FT HELIX 58 60

FT STRAND 62 66

FT TURN 69 70

FT STRAND 71 76

FT TURN 77 80

FT STRAND 81 85

FT TURN 87 88

FT STRAND 91 94

FT HELIX 99 101

FT STRAND 103 107

FT TURN 109 110

FT STRAND 113 117

FT TURN 121 122

FT STRAND 124 124

FT STRAND 127 127

FT TURN 129 130

FT STRAND 132 133

FT HELIX 136 138

FT TURN 141 142

FT HELIX 144 146

FT STRAND 148 152

Query Match 33.1%; Score 60; DB 1; Length 155;

Best Local Similarity 40.0%; Pred. No. 0.21;

Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 3 WCTWDAFYLTVPQGVIEGVRHLVD 27

Db 33 YCKNGEFFLRHPDGRVDGVRKSD 57

RESULT 6

FGF2_SHEEP STANDARD: PRT; 155 AA.

ID P20003;

AC 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast

DE growth factor) (BFGF) (Prostatropin).

GN FGF2 OR FGF-2.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;

OC Bovidae; Caprinae; Ovis.

OX NCBI_TaxID=9940;

RN [1]

RP SEQUENCE FROM N.A.

RA Sutton R., Ward W.G., Raphael K.A., Cam G.R.;

RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE OF 9-155.

RX MEDLINE=88055577; PubMed=3678486;

RA Simpson R.J., Moritz R.L., Lloyd C.J., Fabrl L.J., Nice E.C.,

RA Rubira M.R., Burgess A.W.;

RT "Primary structure of ovine pituitary basic fibroblast growth

RT factor.";

RT FEBS Lett. 224:128-132(1987).

CC -1- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS

CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN

CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND

CC CONCENTRATION OF THESE 2 GROWTH FACTORS.

CC -1- SUBUNIT: MONOMER.

CC -1- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES

CC AFGF.

CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>

CC or send an email to license@isb-sdb.ch).

CC -----

DR EMBL; L36136; AAA31519.1; -.

DR PIR; S00185; S00185.

DR HSP; P09038; 1BFF.

DR InterPro; IPR002209; HB/F-growthfact.

DR InterPro; IPR002348; ILI_HBGF.

DR Pfam; PF00167; FGF; 1.

DR PRINTS; PR00262; ILI_HBGF.

```

DR ProDom: PD000831; HB/F-growthfact; 1.
DR SMART: SM00442; FGF; 1.
DR PROSITE: PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT PROPEP 1 9
FT CHAIN 10 155
FT SITE 45 48
FT SITE 87 90
FT BINDING 27 31
FT BINDING 116 119
SQ SEQUENCE 155 AA; 17280 MW; B5F2364BA610606D CRC64;

Query Match
Best Local Similarity 40.0%; Score 60; DB 1; Length 155;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTMDAFYLVHPGVIEGVRLVD 27
DB 33 YCKNGGFRLRHPDGRVDRKSD 57

RESULT 7
FGF2_MONDO
ID FGF2_MONDO STANDARD: PRT; 156 AA.
AC P48798;
DT 01-FEB-1996 (rel. 33, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast growth factor) (BGF) (Prostatropin).
GN FGF2
OS Monodelphis domestica (Short-tailed grey opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.
RX NCBI_TaxID=13616;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Eye;
RX MEDLINE-94296558; PubMed-8024698;
RA Kuswitt D.F., Sabourin C.L.K., Sherburn T.E., Ley R.D.;
RT "Characterization of cDNA encoding basic fibroblast growth factor of the marsupial Monodelphis domestica.";
RL DNA Cell Biol. 13:549-554(1994).
CC -1- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND CONCENTRATION OF THESE 2 GROWTH FACTORS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES ARGF.
CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z15154; CAAT8854.1; ALT_INIT.
DR HSPF: P09038; 1BRF.
DR InterPro: IPR002209; HB/F-growthfact.
DR InterPro: IPR002348; IIL_HBGF.
DR Pfam: PF00167; FGF; 1.
DR PRINTS: PR00262; IILHBGF.
DR ProDom: PD000831; HB/F-growthfact; 1.
DR SMART: SM00442; FGF; 1.
DR PROSITE: PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT PROPEP 1 9
FT CHAIN 10 156
FT BINDING 28 32
BY SIMILARITY.
HEPARIN-BINDING GROWTH FACTOR 2.
HEPARIN (POTENTIAL).

```

```

FT BINDING 117 120
SQ SEQUENCE 156 AA; 17303 MW; 7E655FCC49BF1209 CRC64;
HEPARIN (POTENTIAL).

Query Match
Best Local Similarity 36.0%; Score 59; DB 1; Length 156;
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTMDAFYLVHPGVIEGVRLVD 27
DB 34 YCKNGGFRLRHPDGRVDRKSD 58

RESULT 8
SSUA_ECOLI
ID SSUA_ECOLI STANDARD: PRT; 319 AA.
AC P75853; O9R705;
DT 15-JUL-1998 (rel. 36, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Putative aliphatic sulfonates binding protein precursor.
GN SSUA OR B0936.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-99436146; PubMed-10506196;
RA Van der Ploeg J.R., Iwanicka-Nowicka R., Bykowski T., Hryniewicz M., Leisinger T.;
RT "The Escherichia coli ssuEADCB gene cluster is required for the utilization of sulfur from aliphatic sulfonates and is regulated by the transcriptional activator Cbl.";
RL J. Biol. Chem. 274:29358-29365(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Alida H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiiuchi T.;
RT "A 718-Kb DNA Sequence of Escherichia coli K-12 Genome Corresponding to the 12.7-28.0 min Region on the Linkage Map.";
RL DNA Res. 3:137-155(1996).
CC -1- FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR ALIPHATIC SULFONATES. PUTATIVE BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: TO B.SUBTILIS SSUA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ237695; CAB40390.1;
DR EMBL: AE000195; AAC74022.1; ALT_INIT.

```

```

DR EMBL; D90732; BAA35691.1; ALT_INIT.
DR EcoGene; EG13707; sauA.
DR InterPro; IPR001638; SBP_bac_3.
DR SMART; SM00062; PBPB; 1.
DR Transprot; Periplasmic; Signal: Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 319 PUTATIVE ALIPHATIC SULFONATES BINDING
FT PROTEIN.
SO SEQUENCE 319 AA; 34557 MW; BACB0711F38D53D0 CRC64;

Query Match 29.6%; Score 53.5; DB 1; Length 319;
Best Local Similarity 46.2%; Pred. No. 3.4;
Matches 12; Conservative 1; Mismatches 10; Indels 3; Gaps 1;

Oy 3 WCTMDAFYLVHPGVIGVRLHYDG 28
Db 180 WAIMDPTYSALLQG--GVRLKDG 202

RESULT 9
FGF2_CHICK STANDARD: PRT: 158 AA.
AC P48800.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (basic fibroblast
DE growth factor) (BFGF).
GN FGF2 OR FGF-2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93246053; PubMed=7683281;
RA Borja A.Z., Zeller R., Meljers C.;
RT "Expression of alternatively spliced bfgf first coding exons and
RT antisense mRNAs during chicken embryogenesis.";
RL Dev. Biol. 157:110-118(1993).
CC -1- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS
CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN
CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND
CC CONCENTRATION OF THESE 2 GROWTH FACTORS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES
CC AREG.
CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95707; AAA48617.1; -.
DR HSSP; P09038; 1BFF.
DR InterPro; IPR002209; HB/F_growthfact.
DR InterPro; IPR002348; IIL_HBGF.
DR Pfam; PF00167; FGF_1.
DR PRINTS; PR00262; IILHBGF.
DR PRODOM; PD000831; HB/F_growthfact. 1.
DR SMART; SM00442; FGF_1.
DR PROSITE; PS00247; HBGF_FGF_1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT PROPEP 1 12 BY SIMILARITY.
FT CHAIN 13 158 HEPARIN-BINDING GROWTH FACTOR 2.
FT BINDING 30 34 HEPARIN (POTENTIAL).
FT BINDING 119 122 HEPARIN (POTENTIAL).
SO SEQUENCE 158 AA; 17374 MW; 7869B84C17F1B16 CRC64;

```

```

Query Match 29.3%; Score 53; DB 1; Length 158;
Best Local Similarity 36.0%; Pred. No. 2;
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Oy 3 WCTMDAFYLVHPGVIGVRLHYDG 27
Db 36 YCKNGFFLRINPDGRVDVREKSD 60

RESULT 10
PTP6_DROME STANDARD: PRT: 1462 AA.
ID PTP6_DROME
AC P16620.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase DPTP precursor (EC 3.1.3.48) (Protein-
DE tyrosine-phosphate phosphohydrolase).
GN PTP69D OR DPTP.
OS Drosophila melanogaster (Fruit fly).
OC Insecta; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Eukaryota; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90045860; PubMed=2554325;
RA Streuli M., Krueger N.X., Tsai A.Y.M., Salto H.;
RT "A family of receptor-linked protein tyrosine phosphatases in humans
RT and Drosophila.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).
CC -1- FUNCTION: IT IS POSSIBLE THAT DPTP IS A CELL ADHESION RECEPTOR
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein
CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27699; AAA28842.1; -.
DR PIR; B36182; B36182.
DR HSSP; P18052; IYFO.
DR FLYBASE; FBgn0014007; Ptp69D.
DR InterPro; IPR003861; FN_III.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003600; IG_C2.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_PP.
DR Pfam; PF00041; fn3_3.
DR Pfam; PF00047; Ig_2.
DR Pfam; PF00102; Y_phosphatase_2.
DR PRINTS; PR00700; PTPYPPHPTASE.
DR SMART; SM00060; FN3_3.
DR SMART; SM00410; IG_1like_1.
DR SMART; SM00408; IGC2_1.
DR SMART; SM00194; PTPc_2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PP; 2.
KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
KW Cell adhesion; Immunoglobulin domain; Repeat.
FT SIGNAL 1 23
FT CHAIN 24 1462 PROTEIN-TYROSINE PHOSPHATASE DPTP.

```

FT DOMAIN 24 805 EXTRACELLULAR (POTENTIAL).
 FT TRASMEN 806 823 POTENTIAL.
 FT DOMAIN 824 1462 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 38 119 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 147 221 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 332 437 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 438 538 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 912 1165 PROTEIN-TYROSINE PHOSPHATASE 1.
 FT DOMAIN 1208 1459 PROTEIN-TYROSINE PHOSPHATASE 2.
 FT ACT_SITE 1097 1097 BY SIMILARITY.
 FT ACT_SITE 1391 1391 BY SIMILARITY.
 FT DISULFID 45 112 POTENTIAL.
 FT DISULFID 154 214 POTENTIAL.
 FT CARBOHYD 58 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 58 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 451 451 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 516 516 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 701 701 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1462 AA; 167411 MW; F8091D69E88230EB CRC64;

Query Match 28.7%; Score 52; DB 1; Length 1462;
 Best Local Similarity 32.0%; Pred. No. 24;
 Matches 8; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Oy 1 FGMCTWDAFLTVHPGVIEGRHL 25
 Db 1056 YHUTWKPMAEPHPIKTRQI 1082

RESULT 11
 H181_ANASP STANDARD: PRT: 353 AA.
 AC Q8YV69; 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Histidinol-phosphate aminotransferase 1 (EC 2.6.1.9) (Imidazole
 DE acetol-phosphate transaminase 1).
 GN HIS1 OR ALR2092.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Muraiki A.,
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 CC -1- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate = 3-
 CC (imidazol-4-yl)-2-oxopropyl phosphate + L-glutamate.
 CC -1- COFACTOR: Pyridoxal phosphate (By similarity).
 CC -1- PATHWAY: Histidine biosynthesis; seventh step.

CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES. HISTIDINOL-PHOSPHATE AMINOTRANSFERASES
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AP003588; BAB73791.1; -
 DR InterPro: IPR004839; AminoTransf1/2.
 DR InterPro: IPR001917; NHTransf.2.
 DR Pfam: PF00155; aminoTran.1.2; 1.
 DR TIGRfam: TIGR01141; hisC.1.
 DR PROSITE: PS00599; AA-TRANSFER_CLASS_2; 1.
 KW Histidine biosynthesis; transferase; aminotransferase;
 KW pyridoxal phosphate; Complete proteome.
 FT BINDING 211 211 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 SQ SEQUENCE 353 AA; 39105 MW; 3891CF87E990A4E CRC64;

Query Match 28.5%; Score 51.5; DB 1; Length 353;
 Best Local Similarity 50.0%; Pred. No. 7.1;
 Matches 11; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

Oy 2 GWCMTWDA---FYLTVHPGVIE 20
 Db 282 GFCWDSQANFLTPPGQNAE 303

RESULT 12
 E312_ADE03 STANDARD: PRT: 106 AA.
 AC P11319; 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE Probable early E3 12.1 kDa glycoprotein.
 DE Human adenovirus type 3.
 OS Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=43659;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87219876; PubMed=3582978;
 RA Signaes C., Akusjaervi G., Pettersson U.;
 RT "Region E3 of human adenoviruses; differences between the oncogenic
 RT adenovirus-3 and the non-oncogenic adenovirus-2.";
 RL Gene 50:173-184(1986).
 CC -1- FUNCTION: NOT YET KNOWN.
 CC -1- SIMILARITY: TO EQUIVALENT PROTEIN IN OTHER ADENOVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M15952; AAA42481.1; -
 DR PIR: B29500; ERAD31.
 KW Early protein; Glycoprotein.
 FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 106 AA; 12124 MW; C0374ACAB6CECEC0 CRC64;

Query Match 28.2%; Score 51; DB 1; Length 106;
 Best Local Similarity 56.2%; Pred. No. 2.6;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 14 HPGGVIEGRHLVDSG 29
 Db 111 111 111 111


```

Db      40  HPGCPAHGVRITIEGG  55

RESULT 13
ID_PIP_NEIGO STANDARD; PRT; 310 AA.
AC P42786;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Proline iminopeptidase (EC 3.4.11.5) (PIP) (Prolyl aminopeptidase)
OS (PAP).
GN PIP.
OS Neisseria gonorrhoeae.
CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
CX NCBI_TaxID=485;
XX [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RP STRAIN=MS11 / MS01-1X;
RC MEDLINE=95020651; PubMed=7934933;
RA Albertson N.H., Koomey M.;
RT "Molecular cloning and characterization of a proline iminopeptidase gene from Neisseria gonorrhoeae."
RL Mol. Microbiol. 9:1203-1211(1993).
CC -1 FUNCTION: HYDROLYSES PEPTIDES HAVING THE STRUCTURE PRO-Y-Z TO YIELD FREE PROLINE. ALSO HYDOLYSES THE DIPEPTIDE PRO-GLY.
CC -1 CATALYTIC ACTIVITY: Release of a N-terminal proline from a peptide.
CC -1 SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S33.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its use by
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/).
CC CC or send an email to license@isb-sdb.ch).
-----
CC EMBL: Z25461; CAA80948.1; -.
DR HSSP: O32449; 1QPR.
DR MEROPS: S33.001; -.
DR InterPro: IPR0000073; Abhydrolase.
DR InterPro: IPR002410; Pro_Aminopeptase.
DR InterPro: IPR000379; Ser_estr_sste.
DR Pfam: PF00561; abhydrolase_1.
DR PRINTS: PR00793; PROAMNOPTASE.
DR TIGRFAMS: TIGR01249; pro_Imino_pep_1; 1.
DR KW Hydrolyase; Aminopeptidase.
FT ACT_SITE 107 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 260 BY SIMILARITY.
FT ACT_SITE 287 PROTON DONOR (BY SIMILARITY).
SO SEQUENCE 310 AA; 34791 MW; D0E7ABEA908A1AEO CRC64;

Query Match 27.3%; Score 49.5; DB 1; Length 310;
Best Local Similarity 23.7%; Pred. No. 12;
Matches 9; Conservative 7; Mismatches 11; Indels 11; Gaps 1;

OY 3 WCTWDAYFLTVHPGVIEGV-----RHLYDGG 29
   1::: : 1:11 1 : 1:11
DB 195 WADWESYLIRFEPGEVDEDAVYASLAIRLENHYFYVNGG 232

RESULT 14
ID_YAED_ECOLI STANDARD; PRT; 191 AA.
AC P31546;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yaed.
OS Escherichia coli, and

```

OS *Escherichia coli* O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteria.
OX NCBI_TaxID=562, 83334;
OX 11)
RP SEQUENCE FROM N.A.
RP STRAIN-K12;
RA Miyamoto K.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RC MEDLINE=97426617, PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RA Takemoto K., Mori T., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RA "Systematic sequencing of the *Escherichia coli* genome: analysis of the
RT 4.0 - 6.0 min (189,987 - 281,415bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RA Ohnishi M., Murata T., Nakayama K., Kuhara S., Hattori M.,
RA Kishikawa K., Yasunaga T., Makino K., Shinagawa H., Hayashi T.;
RT "Development of primer sets for direct sequence determination of all
RL the ribosomal operons of *Escherichia coli*.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935, PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grodbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatter F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
RL Nature 409:529-533(2001).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RC MEDLINE=21156231, PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Saekawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
RT O157:H7 and genetic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -I- SIMILARITY: STRONG, TO H.INFLUENZAE H10621.1.
CC -I- SIMILARITY: TO M.JANNASCHII M00015.
CC -I- SIMILARITY: TO THE N-TERMINAL PART OF E.COLI HISB, THE HISTIDINOL
CC PHOSPHATE PHOSPHATASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D15061; AAC03661.1; ALT_INIT.
 DR EMBL; AE000129; BAA7331.1; -.
 DR EMBL; D83536; BAA77877.1; -.
 DR EMBL; U70214; AAB08628.1; -.
 DR EMBL; AB035926; BAA93568.1; -.
 DR EMBL; AE005195; AAG54502.1; -.
 DR EMBL; AP002550; BAB33625.1; -.
 DR EMBL; EG11736; yaed.
 DR InterPro: IPR001454; Hlg_nase/hydrlase.
 DR InterPro: IPR004446; Hstcd_phs_rel.
 DR Pfam: PF00702; Hydrolase; 1.
 DR TIGRfams; TIGR00213; hstcd_phs_rel; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 191 AA; 21294 MW; E7814B34A23128FA CRC64;
 SQ

Query Match 27.1%; Score 49; DB 1; Length 191;
 Best Local Similarity 28.6%; Pred. No. 8.8;
 Matches 10; Conservative 3; Mismatches 12; Indels 10; Gaps 1;

OY 3 WCTWD-----AFYLVHPGVIEGVRHLVD 27
 DB 74 WMDWSLADRDVDLDGIRYCPHPSVEEFQVCD 108

RESULT 15
 LIPB_SYNY3 STANDARD; PRT; 227 AA.
 AC P74519;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lipote-protein ligase B (EC 6.-.-) (Lipote biosynthesis protein
 B).
 GN LIPB OR SLR0994.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
 RA Hosonouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimpō S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -!- FUNCTION: INVOLVED IN THE ATTACHMENT OF LIPOYL GROUPS TO PROTEINS,
 CC BY CREATING AN AMIDE LINKAGE THAT JOINS THE FREE CARBOXYL GROUP OF
 CC LIPOIC ACID TO THE EPSILON-AMINO GROUP OF A SPECIFIC LYSINE
 CC RESIDUE IN LIPOYLATED PROTEINS (BY SIMILARITY).
 CC -!- PATHWAY: Lipote biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE LIPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D90915; BAA18623.1; -.
 DR InterPro: IPR004143; BPL_LipA_LipB.

DR InterPro: IPR000544; Lipote_B.
 DR Pfam: PF03099; BPL_LipA_LipB; 1.
 DR ProDom: PD006086; Lipote_B; 1.
 DR TIGRfams; TIGR00214; LipB; 1.
 DR PROSITE: PS01313; LipB; 1.
 KW Lipase; Complete proteome.
 SQ SEQUENCE 227 AA; 25640 MW; 0D89E0C781030A46 CRC64;

Query Match 27.1%; Score 49; DB 1; Length 227;
 Best Local Similarity 39.3%; Pred. No. 10;
 Matches 11; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

OY 2 GWCWDAPYLVHPGVIEGVRHLVDG 29
 DB 150 GWVTCGFAFNICPD--LEGFSHIVPCG 175

Search completed: March 27, 2003, 10:04:15
 Job time : 5.93902 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 16.8293 Seconds
(without alignments)
367.301 Million cell updates/sec

Title: US-09-675-208-1

Perfect score: 181
Sequence: 1 FGWCTMDAFYLVHPQGVIEGVRHLVDGCG 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	100.0	784	10	Q92T62 cucumis sat
2	153	84.5	816	10	Q942T6 Oryza sativ
3	147	81.2	798	10	Q8VWV6 Pisum sativ
4	144	79.6	783	10	Q8FVND9 Arabidopsis
5	124	68.5	863	10	Q8RW08 Arabidopsis
6	117	64.6	857	10	Q8W08 Arabidopsis
7	111	61.3	765	10	Q43408 Arabidopsis
8	111	61.3	853	10	Q43408 Arabidopsis
9	108	59.7	757	10	Q39466 Arabidopsis
10	108	59.7	757	10	Q39466 Arabidopsis
11	108	59.7	773	10	Q40077 Arabidopsis
12	108	59.7	773	10	Q40077 Arabidopsis
13	106	58.6	807	10	Q9SYT4 Arabidopsis
14	101	55.8	779	10	Q9M4M7 Arabidopsis
15	99	54.7	844	10	Q8RX87 Arabidopsis
16	82	45.3	1170	10	Q9LFZ7 Arabidopsis

17	72	39.8	371	10	Q04607 Arabidopsis
18	64.5	35.6	374	16	Q9KRP2 Arabidopsis
19	63	34.8	702	2	Q93CM6 Arabidopsis
20	61	33.7	198	16	Q92KR4 Arabidopsis
21	60.5	33.4	674	17	Q96XG2 Arabidopsis
22	60	33.1	114	4	Q16443 Arabidopsis
23	60	33.1	114	4	Q00527 Arabidopsis
24	60	33.1	130	6	Q77767 Arabidopsis
25	60	33.1	153	11	Q925A3 Arabidopsis
26	60	33.1	196	4	P78443 Arabidopsis
27	60	33.1	451	5	Q21801 Arabidopsis
28	57.5	31.8	649	17	Q97094 Arabidopsis
29	57	31.5	170	11	Q60487 Arabidopsis
30	56.5	31.2	221	5	Q9GPI2 Arabidopsis
31	56.5	31.2	373	16	Q82B05 Arabidopsis
32	54	29.8	106	12	Q8UW75 Arabidopsis
33	54	29.8	390	10	Q43811 Arabidopsis
34	54	29.8	390	10	Q43812 Arabidopsis
35	54	29.8	489	16	P93772 Arabidopsis
36	53.5	29.6	319	16	Q8XDP8 Arabidopsis
37	53	29.3	557	16	Q9KEU2 Arabidopsis
38	52.5	29.0	265	2	P72162 Arabidopsis
39	52.5	29.0	265	16	Q9HTB8 Arabidopsis
40	52	28.7	237	10	Q9C7N3 Arabidopsis
41	52	28.7	619	16	Q9K228 Arabidopsis
42	52	28.7	619	16	Q92935 Arabidopsis
43	52	28.7	641	17	Q978X5 Arabidopsis
44	52	28.7	650	4	Q96ED3 Arabidopsis
45	52	28.7	714	4	Q8WU76 Arabidopsis

ALIGNMENTS

RESULT 1
ID Q92T62 PRELIMINARY: PRT: 784 AA.
AC Q92T62;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Raffinose synthase (EC 2.4.1.82).
GN RFS.
OS Cucumis sativus (Cucumber).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBT_TaxID=3659;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Ohsumi C., Nozaki J., Kida T.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF073744; AAD02832.1; -;
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 784 AA: 86920 MW: 3806A91F0908933 CRC64;

Query Match 100.0%; Score 181; DB 10; Length 784;
Best Local Similarity 100.0%; Pred. No. 7e-17;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGWCTMDAFYLVHPQGVIEGVRHLVDGCG 30
DB 215 FGWCTMDAFYLVHPQGVIEGVRHLVDGCG 244

RESULT 2

ID Q942T6 PRELIMINARY: PRT: 816 AA.
AC Q942T6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE	Putative raffinose synthase.
GN	P0583G08.2.
OS	Oryza sativa (Rice).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC	Euhartioideae; Oryzeae; Oryza.
OX	NCBI_TaxID=4530;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-CV. NIPPONBARE;
RA	Sasaki T., Matsumoto T., Yamamoto K.;
RT	"Oryza sativa nipponbare(gas3) genomic DNA, chromosome 1, PAC
RT	clone:P0583G08."
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AP003282; BAB64768.1; -
SO	SEQUENCE 816 AA; 89588 MW; B316DEF3566C5178 CRC64;
Query Match	84.5%; Score 153; DB 10; Length 816;
Best Local Similarity	83.3%; Pred. NO. 6.2e-13;
Matches 25; Conservative 1; Mismatches 4; Indels 0; Gaps 0	
OY	1 FGMCTWDAFYLTVPHPQGVIEGRHLVDGC 30
Dd	227 FGCTWDAFYLTKVRHKGVEGYRRLLADGC 256
RESULT 3	
O8YWN6	PRELIMINARY; PRT; 798 AA.
ID	O8YWN6
AC	O8YWN6:
DT	01-MAR-2002 (TREMBLrel. 20. Created)
DT	01-MAR-2002 (TREMBLrel. 20. Last sequence update)
DT	01-MAR-2002 (TREMBLrel. 20. Last annotation update)
DE	Raffinose synthase (EC 2.4.1.82).
GN	RF5.
OS	Pisum sativum (Garden pea).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosid II; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX	NCBI_TaxID=3888;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-CV. WUNDER VON KELVEDON; TISSUE-SEED:
RA	Peterbauer T., Mach L., Mucha J., Richter A.;
RT	"Molecular characterization of raffinose synthase from pea (Pisum
RT	sativum L.) seeds."
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AJ426475; CAD20127.2; -
KW	transferase; Glycosyltransferase.
SO	SEQUENCE 798 AA; 88717 MW; 8D3FF3BD5BF8617B7 CRC64;
Query Match	81.2%; Score 147; DB 10; Length 798;
Best Local Similarity	80.0%; Pred. NO. 4.2e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0	
OY	1 FGMCTWDAFYLTVPHPQGVIEGRHLVDGC 30
Dd	235 FGCTWDAFYLTKVRHKGVEGYRKALTDGCC 264
RESULT 4	
O9FND9	PRELIMINARY; PRT; 783 AA.
ID	O9FND9
AC	O9FND9:
DT	01-MAR-2001 (TREMBLrel. 16. Created)
DT	01-MAR-2001 (TREMBLrel. 16. Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21. Last annotation update)
DE	Raffinose synthase protein.
GN	MP012.13.
OS	Aralidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.

[illegible]

Db 224 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 252

RESULT 6

09SBZ0 PRELIMINARY: PRT: 857 AA.
 ID 09SBZ0:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Galactinol-4-epimerase galactosyltransferase (EC 2.4.1.67).
 OS Phaseolus angularis (Adzuki bean) (Vigna angularis).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
 CC NCBI_TaxID=3914;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20117502; PubMed=10652123;
 RA Peterbauer T., Mucha J., Mayer U., Popp M., Gloessel J., Richter A.;
 RT "Stachyose synthase in seeds of adzuki bean (Vigna angularis).";
 RT molecular cloning and functional expression of stachyose synthase.";
 RL Plant J. 20:309-318(1999).
 DR EMBL: Y19024; CAB64363.1; -;
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 857 AA; 94949 MW; 85248C4B81165679 CRC64;

Query Match 64.6%; Score 117; DB 10; Length 857;
 Best Local Similarity 65.5%; Pred. No. 7.4e-08;
 Matches 19; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 29
 Db 221 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 249

RESULT 7

043408 PRELIMINARY: PRT: 765 AA.
 ID 043408:
 AC 043408:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Putative imbibition protein.
 OS Brassica oleracea (Cauliflower).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Brassicales; Brassicaceae; Brassica.
 CC NCBI_TaxID=3712;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. ALPHA BALOMA; TISSUE=ROOT TIPS;
 RA Fujikura Y., Karsen C.K.;
 RT "Cauliflower cDNA encoding a putative imbibition protein.";
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X79330; CAA55893.1; -;
 SQ SEQUENCE 765 AA; 84084 MW; 73398603048E9B58 CRC64;

Query Match 61.3%; Score 111; DB 10; Length 765;
 Best Local Similarity 65.5%; Pred. No. 4.5e-07;
 Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 29
 Db 198 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 226

RESULT 8

093XK2 PRELIMINARY: PRT: 853 AA.
 ID 093XK2:
 AC 093XK2:
 DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Stachyose synthase (EC 2.4.1.67).
 GN STSL.

OS Pisum sativum (Garden pea).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 CC NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. WUNDER VON KELYDON; TISSUE=SEED;
 RA Peterbauer T.;
 RT "Molecular cloning of a cDNA encoding for stachyose synthase from pea
 seeds.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ311087; CAC38094.1; -;
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 853 AA; 95890 MW; CB10F18C8D37B38C CRC64;

Query Match 61.3%; Score 111; DB 10; Length 853;
 Best Local Similarity 62.1%; Pred. No. 5.1e-07;
 Matches 18; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 29
 Db 221 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 249

RESULT 9

039466 PRELIMINARY: PRT: 357 AA.
 ID 039466:
 AC 039466:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Seed imbibition protein (Fragment).
 GN SIP.
 OS Cicer arietinum (Chickpea) (Garbanzo).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.
 CC NCBI_TaxID=3827;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cervantes E.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X95875; CAA65125.1; -;
 FT NON_TER 1 1
 FT NON_TER 357 357
 SQ SEQUENCE 357 AA; 39465 MW; E8CD30399E1BF3EE CRC64;

Query Match 59.7%; Score 108; DB 10; Length 357;
 Best Local Similarity 62.1%; Pred. No. 5.2e-07;
 Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 29
 Db 31 FGWCTWDAFYLTVPAGIWMGVKEPSSDGG 59

RESULT 10

040077 PRELIMINARY: PRT: 757 AA.
 ID 040077:
 AC 040077:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Seed imbibition protein.
 GN Sipl.
 OS Hordeum vulgare (Barley).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

OC Triliceae; Hordeum.
 OX NCB1_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HIMALAYA
 RA Heck G.R., Dorsett C., Ho T.-H.D.
 RT "Cloning and characterization of a gene, Slp1, associated with seed
 RL "Imbibition in barley."
 DR EMBL: M77475; AAA32975.1;
 SQ SEQUENCE 757 AA; 82133 MW; EA9E7B771AABCCF CRC64;

Query Match 59.7%; Score 108; DB 10; Length 757;
 Best Local Similarity 62.1%; Pred. No. 1.2e-06;
 Matches 18; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 FGMCTWDATYLVHPQVIEGVRLHVDG 29
 DB 205 FGMCTWDATYLVHPQVIEGVRLHVDG 233

RESULT 11

O9SCM1 PRELIMINARY; PRT; 773 AA.
 AC O9SCM1:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE Imbibition protein homolog.
 GN T8H10.120.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Borkova D., Ansoorge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Quetler F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL13246; CAB66109.1;
 SQ SEQUENCE 773 AA; 85143 MW; 0852F9E67952C8D3 CRC64;

Query Match 59.7%; Score 108; DB 10; Length 773;
 Best Local Similarity 62.1%; Pred. No. 1.2e-06;
 Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGMCTWDATYLVHPQVIEGVRLHVDG 29
 DB 200 FGMCTWDATYLVHPQVIEGVRLHVDG 228

RESULT 12

O94A08 PRELIMINARY; PRT; 773 AA.
 AC O94A08:
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE Putative imbibition protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
 RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,

RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
 RA Caminici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene T8H10.120/AT3g57520 (GI:6706423).";
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY050772; AAK92707.1;
 SQ SEQUENCE 773 AA; 85171 MW; 18BD9E67952C801 CRC64;

Query Match 59.7%; Score 108; DB 10; Length 773;
 Best Local Similarity 62.1%; Pred. No. 1.2e-06;
 Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGMCTWDATYLVHPQVIEGVRLHVDG 29
 DB 200 FGMCTWDATYLVHPQVIEGVRLHVDG 228

RESULT 13

O9SYJ4 PRELIMINARY; PRT; 807 AA.
 AC O9SYJ4:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE Putative raffinose synthase OR seed IMBIBITION protein.
 GN T7B11.23 OR AT4G01970.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA:
 RA Huang E.N., Nascimento L., de la Bastide M., Habermann K., Vil M.D.,
 RA Preston R.R., Spiegel L.A., See L.H., Shah R., Matero A.,
 RA O'Shaughnessy A., Rodriguez M., Shekher M., Swaby I., Schutz K.,
 RA Parnell L.D., Dedhia N.N., McCombie W.R.;
 RT "Arabidopsis thaliana BAC T7B11 from chromosome IV near 10 cM.";
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., Huang E.N., Nascimento L.O., de la Bastide M., Vil D.M.,
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC007138; AAD22659.1;
 DR EMBL: AL161493; CAB80690.1;
 SQ SEQUENCE 807 AA; 90122 MW; 963DCD5A827B338B CRC64;

Query Match 58.6%; Score 106; DB 10; Length 807;
 Best Local Similarity 62.1%; Pred. No. 2.4e-06;
 Matches 18; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 1 FGMCTWDATYLVHPQVIEGVRLHVDG 29
 DB 235 FGMCTWDATYLVHPQVIEGVRLHVDG 263

RESULT 14

O9M4M7 PRELIMINARY; PRT; 779 AA.
 AC O9M4M7:
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Putative seed imbibition protein.
 GN SIP.
 OS Persa americana (Avocado).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
 OX NCBI_TaxID=3435;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, HASS; TISSUE=MESOCARP;
 RA Zamorano J.P., Evans A.D., Doplico B., Lowe A.L., Wilson I.D.,
 RA Merodio C., Grieson D.;
 RT "Isolation and characterization of cDNAs for mRNAs regulated during
 cold storage of avocado (Persea americana Mill.) fruit."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ133148; CAB77245.1;
 SO SEQUENCE 779 AA; 85368 MW; C3A8B43160316785 CRC64;

Query Match 55.8%; Score 101; DB 10; Length 779;
 Best Local Similarity 58.6%; Pred. No. 1.2e-05;
 Matches 17; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 FGMCTMDAFYLTVPQGVTEGVRLVDGG 29
 |||||
 DB 203 FGMCTMDAFETDVTDEGVTEGLKLSGGG 231

RESULT 15

O8RX87 PRELIMINARY; PRT; 844 AA.
 AC O8RX87.
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE AT5G20250/F5024.140.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shun P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Tortum M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY090237; AAL0901.1;
 SO SEQUENCE 844 AA; 94215 MW; 2AC8AB0EA43F8056 CRC64;

Query Match 54.7%; Score 99; DB 10; Length 844;
 Best Local Similarity 58.6%; Pred. No. 2.4e-05;
 Matches 17; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 FGMCTMDAFYLTVPQGVTEGVRLVDGG 29
 |||||
 DB 301 FGMCTMDAFYQEVTOGVTEGLKLSLACG 329

Search completed: March 27, 2003, 10:05:56
 Job time : 20.8293 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:04:25 ; Search time 6.95122 seconds
(without alignments)
253.439 Million cell updates/sec

Title: US-09-675-208-1
Perfect score: 181
Sequence: 1 FGMCTWDAFLTVHPQGVIEGVRLVDDGC 30

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCR_NEW_PUB pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCRUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	33.1	134	9 US-09-901-938-24	Sequence 24, Appl1
2	60	33.1	145	9 US-09-940-601-3	Sequence 3, Appl1
3	60	33.1	146	9 US-10-131-965-3	Sequence 3, Appl1
4	60	33.1	146	9 US-10-131-965-5	Sequence 5, Appl1
5	60	33.1	146	9 US-10-108-195-2	Sequence 2, Appl1
6	60	33.1	146	9 US-10-108-195-3	Sequence 3, Appl1
7	60	33.1	146	9 US-10-108-195-4	Sequence 4, Appl1
8	60	33.1	146	9 US-10-108-195-7	Sequence 7, Appl1
9	60	33.1	146	10 US-09-802-365-2	Sequence 2, Appl1
10	60	33.1	146	10 US-09-802-365-4	Sequence 2, Appl1
11	60	33.1	146	10 US-09-771-302-2	Sequence 2, Appl1
12	60	33.1	146	10 US-09-886-856-2	Sequence 2, Appl1
13	60	33.1	146	10 US-09-886-856-4	Sequence 8, Appl1
14	60	33.1	147	9 US-09-820-596-8	Sequence 8, Appl1
15	60	33.1	150	12 US-10-016-447-8	Sequence 8, Appl1
16	60	33.1	155	9 US-10-081-347-28	Sequence 28, Appl1
17	60	33.1	155	9 US-10-108-195-1	Sequence 1, Appl1
18	60	33.1	155	9 US-09-902-460-2	Sequence 2, Appl1
19	60	33.1	155	10 US-09-822-485-5	Sequence 5, Appl1

20	60	33.1	155	10 US-09-802-365-6	Sequence 6, Appl1
21	60	33.1	155	10 US-09-802-365-8	Sequence 8, Appl1
22	60	33.1	155	10 US-09-251-263-10	Sequence 10, Appl1
23	60	33.1	155	10 US-09-425-021-10	Sequence 10, Appl1
24	60	33.1	155	10 US-09-886-856-6	Sequence 6, Appl1
25	60	33.1	155	10 US-09-886-856-8	Sequence 8, Appl1
26	60	33.1	155	10 US-09-749-728B-7	Sequence 7, Appl1
27	60	33.1	158	10 US-09-826-210-2	Sequence 2, Appl1
28	60	33.1	159	10 US-09-934-706-2	Sequence 2, Appl1
29	60	33.1	210	10 US-09-902-773A-4	Sequence 4, Appl1
30	60	33.1	333	9 US-10-108-195-6	Sequence 6, Appl1
31	60	33.1	501	10 US-09-934-706-4	Sequence 4, Appl1
32	53.5	29.6	333	10 US-09-815-242-10116	Sequence 10116, A
33	52	28.7	313	10 US-09-788-626-19	Sequence 19, Appl1
34	51.5	28.5	280	10 US-09-815-242-11714	Sequence 11714, A
35	49	27.1	140	9 US-10-131-965-2	Sequence 2, Appl1
36	48.5	26.8	161	10 US-09-893-737-288	Sequence 288, App
37	48.5	26.8	497	9 US-09-896-896A-1	Sequence 1, Appl1
38	48.5	26.8	497	10 US-09-753-126-1	Sequence 1, Appl1
39	48.5	26.8	592	10 US-09-753-126-4	Sequence 4, Appl1
40	48	26.5	141	9 US-09-901-938-18	Sequence 18, Appl1
41	48	26.5	207	10 US-09-822-485-19	Sequence 19, Appl1
42	48	26.5	207	10 US-09-750-963-6	Sequence 6, Appl1
43	48	26.5	207	10 US-09-805-805-5	Sequence 5, Appl1
44	48	26.5	207	10 US-09-805-805-7	Sequence 7, Appl1
45	48	26.5	207	10 US-09-817-814-24	Sequence 24, Appl1

ALIGNMENTS

RESULT 1
US-09-901-938-24
Sequence 24, Application US/09901938
Patent No. US20020156001A1
GENERAL INFORMATION:
APPLICANT: ECONS, Michael
APPLICANT: WHITE, Kenneth
APPLICANT: STROM, Tim
APPLICANT: MEITNER, Thomas
TITLE OF INVENTION: NOVEL FIBROBLAST GROWTH FACTOR (FGF23) AND METHODS FOR USE
FILE REFERENCE: 035884-5001
CURRENT APPLICATION NUMBER: US/09/901,938
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/219,137
PRIOR FILING DATE: 2000-07-19
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 24
LENGTH: 134
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-901-938-24

Query Match 33.1%; Score 60; DB 9; Length 134;
Best Local Similarity 40.0%; Pred. No. 0.5;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

CY 3 WCTWDAFLTVHPQGVIEGVRLVDD 27
DB 12 YCKNGCFELRHDPGVRVDSKSD 36

RESULT 2
US-09-940-601-3
Sequence 3, Application US/09940601
Publication No. US20030004319A1
GENERAL INFORMATION:
APPLICANT: Basilio, Claudio
APPLICANT: Dell'Bovi, Pasquale
TITLE OF INVENTION: MAMMALIAN GROWTH FACTOR
FILE REFERENCE: 5986/13586-USC
CURRENT APPLICATION NUMBER: US/09/940,601

```
;; CURRENT FILING DATE: 2001-08-27
;; PRIOR APPLICATION NUMBER: 08/775,567
;; PRIOR FILING DATE: 1996-12-31
;; PRIOR APPLICATION NUMBER: 08/056,482
;; PRIOR FILING DATE: 1993-05-03
;; PRIOR APPLICATION NUMBER: 07/806,771
;; PRIOR FILING DATE: 1991-12-06
;; PRIOR APPLICATION NUMBER: 07/177,506
;; PRIOR FILING DATE: 1988-04-04
;; PRIOR APPLICATION NUMBER: 07/062,925
;; PRIOR FILING DATE: 1987-06-16
;; NUMBER OF SEQ ID NOS: 8
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO: 3
;; LENGTH: 145
;; TYPE: PRP
;; ORGANISM: Bos taurus
US-09-940-601-3
```

```
Query Match          33.1%; Score 60; DB 9; Length 145;
Best Local Similarity 40.0%; Pred. No. 0.54;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
OY      3 MCTWDAFYLTVHPQVIEGVRLVD 27
      :|:|:|:|:|:|:|:|:|:|:|:|
Db      24 YCKNGGFRLRHPDGRVDSVREKSD 48
```

```
RESULT 3
US-10-131-965-3
;; Sequence 3, Application US/10131965
;; Patent No. US20020165160A1
;; GENERAL INFORMATION:
;; APPLICANT: Kavanaugh, Martha J.
;; APPLICANT: Kavanaugh, Michael W.
;; TITLE OF INVENTION: Antigenically Effective Unit Dose of FGF and Method of
;; FILE REFERENCE: 1296/121690505
;; CURRENT APPLICATION NUMBER: US/10/131,965
;; PRIOR FILING DATE: 2002-04-25
;; PRIOR APPLICATION NUMBER: US/09/417,721
;; PRIOR FILING DATE: 1999-10-13
;; PRIOR APPLICATION NUMBER: 60/104,103
;; PRIOR FILING DATE: 1998-10-13
;; NUMBER OF SEQ ID NOS: 15
;; SOFTWARE: Patentl Ver. 2.0
;; SEQ ID NO: 3
;; LENGTH: 146
;; TYPE: PRP
;; ORGANISM: Human FGF-2
US-10-131-965-3
```

```
Query Match          33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
OY      3 MCTWDAFYLTVHPQVIEGVRLVD 27
      :|:|:|:|:|:|:|:|:|:|:|:|
Db      24 YCKNGGFRLRHPDGRVDSVREKSD 48
```

```
RESULT 4
US-10-131-965-5
;; Sequence 5, Application US/10131965
;; Patent No. US20020165160A1
;; GENERAL INFORMATION:
;; APPLICANT: Kavanaugh, Martha J.
;; APPLICANT: Kavanaugh, Michael W.
;; TITLE OF INVENTION: Antigenically Effective Unit Dose of FGF and Method of
;; FILE REFERENCE: 1296/121690505
;; CURRENT APPLICATION NUMBER: US/10/131,965
;; CURRENT FILING DATE: 2002-04-25
```

```
;; PRIOR APPLICATION NUMBER: US/09/417,721
;; PRIOR FILING DATE: 1999-10-13
;; PRIOR APPLICATION NUMBER: 60/104,103
;; PRIOR FILING DATE: 1998-10-13
;; NUMBER OF SEQ ID NOS: 15
;; SOFTWARE: Patentl Ver. 2.0
;; SEQ ID NO: 5
;; LENGTH: 146
;; TYPE: PRP
;; ORGANISM: bovine FGF-2
US-10-131-965-5
```

```
Query Match          33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
OY      3 MCTWDAFYLTVHPQVIEGVRLVD 27
      :|:|:|:|:|:|:|:|:|:|:|:|
Db      24 YCKNGGFRLRHPDGRVDSVREKSD 48
```

```
RESULT 5
US-10-108-195-2
;; Sequence 2, Application US/10108195
;; Publication No. US20030008820A1
;; GENERAL INFORMATION:
;; APPLICANT: Kwan, Chi-Pong
;; APPLICANT: Venkataraman, Ganesh
;; APPLICANT: Shriver, Zachary
;; APPLICANT: Raman, Rahul
;; APPLICANT: Sasisekharan, Ram
;; TITLE OF INVENTION: Methods and Products Related to FGF Dimerization
;; FILE REFERENCE: M00656/70076
;; CURRENT APPLICATION NUMBER: US/10/108,195
;; PRIOR FILING DATE: 2002-03-27
;; PRIOR APPLICATION NUMBER: US 60/279,165
;; PRIOR FILING DATE: 2001-03-27
;; NUMBER OF SEQ ID NOS: 15
;; SOFTWARE: Patentl version 3.1
;; SEQ ID NO: 2
;; LENGTH: 146
;; TYPE: PRP
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Mutant of Native FGF2 with 9 N-terminal Residues Deleted
US-10-108-195-2
```

```
Query Match          33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
OY      3 MCTWDAFYLTVHPQVIEGVRLVD 27
      :|:|:|:|:|:|:|:|:|:|:|:|
Db      24 YCKNGGFRLRHPDGRVDSVREKSD 48
```

```
RESULT 6
US-10-108-195-3
;; Sequence 3, Application US/10108195
;; Publication No. US20030008820A1
;; GENERAL INFORMATION:
;; APPLICANT: Kwan, Chi-Pong
;; APPLICANT: Venkataraman, Ganesh
;; APPLICANT: Shriver, Zachary
;; APPLICANT: Raman, Rahul
;; APPLICANT: Sasisekharan, Ram
;; TITLE OF INVENTION: Methods and Products Related to FGF Dimerization
;; FILE REFERENCE: M00656/70076
;; CURRENT APPLICATION NUMBER: US/10/108,195
;; CURRENT FILING DATE: 2002-03-27
;; PRIOR APPLICATION NUMBER: US 60/279,165
;; PRIOR FILING DATE: 2001-03-27
;; NUMBER OF SEQ ID NOS: 15
```

```

      TYPE: PRT
      ORGANISM: Artificial Sequence
      FEATURE:
      OTHER INFORMATION: Native FGF2 with 9 N-terminal Residues Deleted
      US-108-195-7

Query Match          33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

      QY      3 WCTWDAPFLTVHPQGVIEGRHLVD 27
      :|:|:|:|:|:|:|:|:|
      Db      24 YCKNGCFPLRIHPDGRVDGVREKSD 48

RESULT 9
US-09-802-365-2
; Sequence 2, Application US/09802365
; Patent No. US20020032153A1
; GENERAL INFORMATION:
; APPLICANT: Whitehouse, Martha Jo
; TITLE OF INVENTION: Treatment and Prevention of Erectile Dysfunction
; FILE REFERENCE: 1671.003
; CURRENT APPLICATION NUMBER: US/09/802,365
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,480
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/203,415
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-802-365-2

Query Match          33.1%; Score 60; DB 10; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

      QY      3 WCTWDAPFLTVHPQGVIEGRHLVD 27
      :|:|:|:|:|:|:|:|:|
      Db      24 YCKNGCFPLRIHPDGRVDGVREKSD 48

RESULT 10
US-09-802-365-4
; Sequence 4, Application US/09802365
; Patent No. US20020032153A1
; GENERAL INFORMATION:
; APPLICANT: Whitehouse, Martha Jo
; TITLE OF INVENTION: Methods and Compositions for the
; FILE REFERENCE: 1671.003
; CURRENT APPLICATION NUMBER: US/09/802,365
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,480
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 60/203,415
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-365-4

Query Match          33.1%; Score 60; DB 10; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

```

```

OY      3 WCTWDAFYLLTVHPQGVIEGVRLVD 27
          :| | :| | | :| | | |
Db     24 YCKNGGFLLRIHPDGRVDGVREKSD 48

```

```

? RESULT 11
? US-09-771-302-2
? Sequence 2, Application US/09771302
? Patent No. US20020072489A1
? GENERAL INFORMATION:
? APPLICANT: Whitehouse, Martha J.
? TITLE OF INVENTION: Angiogenetically Effective Unit Dose of FGF-2 and Method
? TITLE OR INVENTION: of Use
? FILE REFERENCE: 1296/12169US04
? CURRENT APPLICATION NUMBER: US/09/771,302
? CURRENT FILING DATE: 2001-01-26
? PRIOR APPLICATION NUMBER: 09/385,114
? PRIOR FILING DATE: 1999-08-27
? PRIOR APPLICATION NUMBER: 60/104,102
? PRIOR FILING DATE: 1998-10-13
? NUMBER OF SEQ ID NOS: 3
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 2
? LENGTH: 146
? TYPE: PRT
? ORGANISM: Bovis bovinus
? US-09-771-302-2

```

```
QY      3 WCITDAFYLLTVHPQGVIEGVRHLVD 27
      : | : | : | : | : | : |
Db     24 YCKNGGFELRIHPDGRVDGVREKSD 48
```

```

RESULT 12
US-09-886-856-2
; Sequence 2, Application US/09886856
; Patent No. US20020115603A1
; GENERAL INFORMATION:
; APPLICANT: Whitehouse, Martha Jo
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Treatment of Peripheral Artery Disease
; FILE REFERENCE: P16090_004
; CURRENT APPLICATION NUMBER: US/09/886,856
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: 60/213,504
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/264,572
; PRIOR FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: 60/276,549
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 146
; TYPE: PRF
; ORGANISM: Bos taurus
; US-09-886-856-2

```

QY 3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
 : | : | : | : | : |
Db 24 YCKNGGFRLRHPDGRVDGVREKSD 48

RESULT 13

```

US-09-886-856-4
? Sequence 4, Application US/09886856
? Patent No. US20020115603A1
? GENERAL INFORMATION:
? APPLICANT: Whitehouse, Martha Jo
? TITLE OF INVENTION: Methods and Compositions for the
? TITL OF INVENTION: Treatment of Peripheral
? FILE REFERENCE: PPI6090.004
? CURRENT APPLICATION NUMBER: US/09/886, 856
? CURRENT FILING DATE: 2001-06-21
? PRIOR APPLICATION NUMBER: 60/213,504
? PRIOR FILING DATE: 2000-06-22
? PRIOR APPLICATION NUMBER: 60/264,572
? PRIOR FILING DATE: 2000-01-26
? PRIOR APPLICATION NUMBER: 60/276,549
? PRIOR FILING DATE: 2001-03-16
? NUMBER OF SEQ ID NOS: 9
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 4
? LENGTH: 146
? TYPE: PRN
? ORGANISM: Homo sapiens
US-09-886-856-4

```

```

QY      3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
      : | : | : | : | : |
Db     24 YCKNGGFRLRHPDGRVDGVREKSD 48

```

```

RESULT 14
US-09-820-596-8
: Sequence 8, Application US/09820596
: Publication No. US2003002170A1
: GENERAL INFORMATION:
: APPLICANT: Khodadoust, Mehran Mohammad
: TITLE OF INVENTION: NOVEL FIBROBLAST GROWTH FACTORS AND
: THERAPEUTIC AND DIAGNOSTIC USES THEREFOR
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: FOLEY, HOAG & ELIOT LLP
: STREET: One Post Office Square
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109-2170
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/820,596
: FILING DATE: 29-Mar-2001
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 09/036,594
: FILING DATE: <unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Arnold E., Beth
: REGISTRATION NUMBER: 35,430
: REFERENCE/DOCKET NUMBER: MIA-026.01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-832-1000
: TELEFAX: 617-832-7000
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 147 amino acids
: TYPE: amino acid
: STRANDEDNESS: <unknown>
: TOPOLOGY: linear

```

```

;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-820-596-8

```

Query Match	33.18;	Score 60;	DB 9;	Length 147;
Best Local Similarity	40.08;	Pred. NO. 0.55;		
Matches 10; Conservative	5;	Mismatches 10;	Indels 0;	Gaps 0;

```

Oy      3 WCTWDAFYLLTVHPQGVIEGVRHLVD 27
          :|  | :| | | :| | | |
Db     25 YCKNGGFLLRIHPDGRVDGVREKSD 49

```

```

RESULT 15
US-10-016-447-8
Sequence 8, Application US/10016447
Patent No. US20020090651A1
GENERAL INFORMATION:
APPLICANT: Kirschner, Marc W.
APPLICANT: Kirschner, Marc W.
TITLE OF INVENTION: Receptor-Ligand Assay
FILE REFERENCE: HU95-01A2
CURRENT APPLICATION NUMBER: US/10/016,447
CURRENT FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/776,207
PRIOR FILING DATE: EARLIER FILING DATE: 1997-06-23
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/441,629
PRIOR FILING DATE: EARLIER FILING DATE: 1995-05-15
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/279,217
PRIOR FILING DATE: EARLIER FILING DATE: 1994-07-22
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 150
TYPE: PR1
ORGANISM: Homo sapien
US-10-016-447-8

```

Query Match	33.1%	Score 60;	DB 12;	Length 150;
Best Local Similarity	40.0%;	Pred. No. 0.56;		
Matches 10;	Conservative 5;	Mismatches 10;	Indels 0;	Gaps 0

```

QY      3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
      .  :| :| :| :| :| :| :| :| :|
Db     33 YCKNGGFFLRHPDGRVDGVREKSD 57

```

Search completed: March 27, 2003, 10:20:35
Job time : 6.95122 secs

